Logica Epistemica Dinamica e Teoria Formale dell'Indagine

PROEMIO

La filosofia analitica del Novecento ha partorito, nel corso dei suoi decenni centrali, due tradizioni di ricerca formale che, pur muovendo da motivazioni teoriche profondamente eterogenee, convergono su un medesimo nucleo problematico: la struttura matematicamente precisa della conoscenza e il suo mutamento ordinato nel tempo. La logica epistemica dinamica — DEL nell'acronimo anglosassone ormai invalso anche nel dibattito continentale — e la teoria formale dell'indagine affidabile nella versione elaborata da Kevin Kelly si propongono entrambe, con strumenti tecnici distinti ma in certi punti sorprendentemente omologhi, di rispondere alla domanda fondamentale su come agenti razionali acquisiscono, rivedono e trasmettono informazione epistemica in risposta a stimoli provenienti dall'ambiente.

Conviene avvertire il lettore fin dall'esordio che le due imprese, pur accostate in questa trattazione, non sono in rapporto di mera contiguità disciplinare. La DEL nasce nell'alveo della logica modale e della semantica dei mondi possibili, nutre i suoi problemi canonici di giochi epistemici, di annunci pubblici e di aggiornamento doxastico, e trova la sua espressione più matura in una serie di contributi che si snoda da Jan Plaza (1989) attraverso Jelle Gerbrandy fino all'opera sistematica di Hans van Ditmarsch, Wiebe van der Hoek e Barteld Kooi. La teoria di Kelly, per converso, affonda le proprie radici nell'epistemologia della scienza, nella teoria della computabilità e in una concezione topologica dello spazio delle ipotesi, e muove da domande che appartengono alla tradizione humiana: in quali circostanze è razionalmente giustificato adottare un metodo d'indagine piuttosto che un altro? Quando un metodo è affidabile? Può il rasoio di Occam ricevere una sanzione puramente formale?

Il punto di contatto più profondo tra le due tradizioni risiede in ciò che si potrebbe chiamare la dinamica della revisione epistemica: entrambe le teorie studiano non stati epistemici statici ma traiettorie — sequenze di stati che si trasformano sotto la pressione di evidenza nuova, di comunicazione, di aggiornamento. La DEL si occupa prevalentemente del momento sincronico della trasformazione: dati uno stato epistemico collettivo e un evento comunicativo (un annuncio, un'azione osservabile o parzialmente osservabile), come si trasforma lo stato? La teoria di Kelly si occupa invece della dimensione diacronica: lungo una sequenza potenzialmente infinita di osservazioni, un agente che adotta un certo metodo d'indagine convergerà alla verità? Con quale efficienza, misurata in termini di cambiamenti di congettura?

La struttura di questa monografia riflette la duplice natura dell'oggetto. La prima parte è interamente dedicata alla DEL: si parte dai fondamenti hintikkaniani della logica epistemica classica, si risale poi all'invenzione plaziana della logica degli annunci pubblici (Public Announcement Logic, PAL), si percorre la tradizione gerbrandyana delle comunicazioni private e dei modelli d'azione, per approdare infine alla sintesi di van Ditmarsch, van der Hoek e Kooi. La seconda parte espone con altrettanta cura la teoria formale dell'apprendimento nella versione kellyiana: si introduce il problema dell'induzione nel suo inquadramento computazionale (Gold 1967), si presenta il framework dell'indagine affidabile, si discute la giustificazione topologica del rasoio di Occam, e si esaminano i risultati principali di caratterizzazione e complessità.

L'apparato matematico è trattato con la massima precisione che l'esposizione prosa consente: tutte le definizioni sono esplicate nella loro forma simbolica, le dimostrazioni di rilievo sono condotte per intero o in forma di abbozzo rigoroso, e si evita accuratamente di invocare risultati non dimostrati o difficilmente reperibili nella letteratura. Il registro adottato è quello della ricerca accademica, non della divulgazione.

Capitolo I — Fondamenti: La Logica Epistemica Classica

I.1 Il punto di partenza hintikkaniano

La logica epistemica nasce, nella sua forma moderna e matematicamente rigorosa, con il volume *Knowledge and Belief* di Jaakko Hintikka, pubblicato nel 1962. L'impresa hintikkaniana consiste nel trattare la conoscenza e la credenza come operatori modali, soggetti a una semantica dei mondi possibili e a un sistema assiomatico derivato dalla logica modale standard. Il punto di partenza filosofico è di matrice kantiana e husserliana: conoscere una proposizione non è un fatto bruto ma una relazione intenzionale strutturata, che può essere analizzata mediante le sue condizioni formali di verità. Hintikka traduce questo intuito in un linguaggio formale preciso, dotato di operatori K (per la conoscenza) e B (per la credenza), e di una semantica relazionale che determina le condizioni di soddisfacibilità di enunciati del tipo "l'agente *i* sa che φ" e "l'agente *i* crede che φ".

La mossa tecnica cruciale è l'adozione della semantica dei mondi possibili nel senso di Kripke, sviluppata negli stessi anni in parallelo e in parte indipendentemente. Un agente cognitivo non conosce, in un dato momento, quale sia il mondo effettivo tra tutti i mondi che egli giudica compatibili con ciò che percepisce e ricorda: la sua relazione di accessibilità epistemica codifica precisamente questa ignoranza residua. L'agente *sa* che φ in un mondo *w* se e soltanto se φ è vera in tutti i mondi che egli ritiene possibili da *w*. Questo schema semantico, nella sua semplicità, ha una potenza espressiva straordinaria: permette di rendere conto dell'ignoranza, della conoscenza

parziale, della conoscenza di secondo ordine (sapere di sapere, sapere di non sapere), e delle strutture epistemiche distribuite su più agenti.

I.2 Il linguaggio formale della logica epistemica multi-agente

Si fissa un insieme finito di agenti A = {1, 2, …, n} e un insieme numerabile Prop = {p, q, r, …} di variabili proposizionali. Il linguaggio della logica epistemica multi-agente, che denoteremo $\mathcal{L}$_EL, è definito dalla seguente grammatica in forma di Backus-Naur:

$$\varphi ::= p \mid \bot \mid \neg\varphi \mid (\varphi \wedge \psi) \mid K_i \varphi$$

dove $p \in$ Prop e $i \in$ A. I connettivi booleani restanti — disgiunzione $\vee$, implicazione $\rightarrow$, equivalenza $\leftrightarrow$, e la costante $\top$ — si definiscono nelle solite maniere abbreviative: $\varphi \vee \psi := \neg(\neg\varphi \wedge \neg\psi)$, $\varphi \rightarrow \psi := \neg\varphi \vee \psi$, $\top := \neg\bot$. L'operatore duale di K_i è M_i ("l'agente i considera possibile che"), definito come $M_i \varphi := \neg K_i \neg\varphi$. Si legge K_i φ come "l'agente i sa che φ" e M_i φ come "l'agente i considera possibile che φ" — o, più tecnicamente, come "φ è epistemicamente compatibile con ciò che l'agente i sa".

La profondità modale di una formula $\varphi \in \mathcal{L}$_EL, indicata md(φ), è definita ricorsivamente:

$$\begin{aligned} md(p) &= 0 \\ md(\bot) &= 0 \\ md(\neg\varphi) &= md(\varphi) \\ md(\varphi \wedge \psi) &= \max(md(\varphi), md(\psi)) \\ md(K_i \varphi) &= 1 + md(\varphi) \end{aligned}$$

La profondità modale misura il grado di annidamento degli operatori epistemici e avrà rilievo tecnico nella dimostrazione del teorema di riduzione per la logica degli annunci pubblici.

I.3 La semantica relazionale di Kripke

Un *modello di Kripke per la logica epistemica* su A e Prop è una tripla

$$M = (W, \{R_i\}_{i \in A}, V)$$

dove W è un insieme non vuoto di *mondi possibili*, per ogni $i \in$ A la relazione $R_i \subseteq W \times W$ è la *relazione di accessibilità epistemica* dell'agente i, e $V : \text{Prop} \rightarrow \mathcal{P}(W)$ è la *funzione di valutazione*, che assegna a ogni variabile proposizionale l'insieme dei mondi in cui essa è vera.

La relazione di soddisfazione M, w $\vDash$ φ, che si legge "φ è vera nel mondo w del modello M", è definita per induzione sulla struttura di φ:

M, w $\vDash$ p iff w $\in$ V(p)
M, w $\vDash$ $\bot$ mai (per nessun w)
M, w $\vDash$ ¬φ iff M, w $\nvDash$ φ
M, w $\vDash$ φ $\wedge$ ψ iff M, w $\vDash$ φ e M, w $\vDash$ ψ
M, w $\vDash$ K_i φ iff per ogni v $\in$ W tale che wR_i v, M, v $\vDash$ φ

La clausola per K_i esprime l'intuizione hintikkaniana: l'agente i sa φ nel mondo w se e soltanto se φ è vera in tutti i mondi che i considera epistemicamente accessibili da w. Una formula φ è *valida* in un modello M, scritto M $\vDash$ φ, se M, w $\vDash$ φ per ogni w $\in$ W. È *valida globalmente* (o semplicemente valida), scritto $\vDash$ φ, se lo è in ogni modello Kripkiano.

I.4 Classi di modelli e sistemi assiomatici

Il contenuto epistemico del sistema logico dipende crucialmente dalle proprietà strutturali imposte sulle relazioni di accessibilità. Proprietà diverse delle R_i corrispondono ad assiomi modali distinti, e dunque a teorie epistemiche con diverso potere espressivo e diversi impegni filosofici.

La corrispondenza tra proprietà relazionali e schemi assiomatici è qui di seguito esposta in dettaglio. Si usa la convenzione di abbreviare "per ogni i $\in$ A" con la locuzione "per ogni agente".

Riflessività ($\forall$ w $\in$ W : wR_i w) corrisponde all'assioma

(T) K_i φ → φ

noto come assioma di *verità* o di *fatticità*. Esso afferma che ciò che un agente sa è vero: la conoscenza, a differenza della credenza, implica la verità del suo contenuto. Questo assioma distingue formalmente K dalla credenza B, che non lo soddisfa in generale.

Transitività ($\forall$ u,v,w : uR_i v $\wedge$ vR_i w → uR_i w) corrisponde all'assioma

(4) K_i φ → K_i K_i φ

l'*introspezione positiva*: se l'agente sa qualcosa, sa di saperlo. Filosoficamente, questo equivale ad affermare che la conoscenza è trasparente al soggetto che la possiede.

Proprietà euclidea ($\forall$ u,v,w : uR_i v $\wedge$ uR_i w → vR_i w) corrisponde all'assioma

(5) ¬K_i φ → K_i¬K_i φ

l'*introspezione negativa*: se l'agente non sa qualcosa, sa di non saperlo. Insieme alla riflessività, la proprietà euclidea implica la simmetria e dunque l'equivalenza, sicché i modelli S5 sono precisamente quelli in cui ogni R_i è una relazione di equivalenza su W.

Serialità ($\forall w \in W : \exists v \in W$ tale che $wR_i v$) corrisponde all'assioma

(D) $\quad K_i \varphi \to M_i \varphi$

che esprime la *consistenza* doxastica: un agente non può credere sia φ che $\neg\varphi$.

Il sistema S5 — ottenuto aggiungendo alla logica proposizionale classica gli assiomi K (distributività: $K_i(\varphi \to \psi) \to (K_i \varphi \to K_i \psi)$), T, 4, 5 e la regola di necessitazione (se $\vdash \varphi$ allora $\vdash K_i \varphi$) — è il sistema assiomatico standard per la conoscenza nel senso epistemico pieno. Esso è suono e completo rispetto alla classe di tutti i modelli di Kripke in cui ogni R_i è una relazione di equivalenza. La completezza di S5 è un risultato classico della logica modale, dimostrabile mediante tecniche di modelli canonici (canonical model construction) oppure per via di tableaux analitici.

Teorema I.1 (Correttezza e completezza di S5): Sia $\varphi \in \mathcal{L}_{EL}$. Allora $\vdash_{S5} \varphi$ se e soltanto se φ è valida in ogni modello di Kripke in cui tutte le relazioni R_i sono relazioni di equivalenza.

La dimostrazione della correttezza ($\vdash_{S5} \varphi \Rightarrow \vDash \varphi$) procede per induzione sulla lunghezza delle derivazioni: si verifica che ogni assioma è valido nella classe intesa e che le regole di inferenza preservano la validità. Per la completezza ($\vDash \varphi \Rightarrow \vdash_{S5} \varphi$) si costruisce il modello canonico $M^c = (W^c, \{R_i^c\}_{i \in A}, V^c)$ in cui:

$W^c = \{ \Gamma \subseteq \mathcal{L}_{EL} : \Gamma \text{ è massimamente consistente in S5} \}$
$\Gamma R_i^c \Delta$ iff $\{ \varphi : K_i \varphi \in \Gamma \} \subseteq \Delta$
$V^c(p) = \{ \Gamma \in W^c : p \in \Gamma \}$

Si dimostra che $M^c, \Gamma \vDash \varphi$ iff $\varphi \in \Gamma$ (lemma di verità) per induzione su φ, e che ogni relazione R_i^c è un'equivalenza grazie agli assiomi 4 e 5. Se φ non è derivabile, allora $\{\neg\varphi\}$ è consistente e si estende a un insieme massimamente consistente $\Gamma \in W^c$ con $\neg\varphi \in \Gamma$, donde $M^c, \Gamma \nvDash \varphi$, contraddicendo l'ipotesi di validità.

I.5 Conoscenza di gruppo: conoscenza comune e conoscenza distribuita

La logica epistemica multi-agente diventa particolarmente interessante quando si considerano operatori collettivi che vanno oltre la semplice iterazione delle conoscenze individuali. Per un insieme non vuoto di agenti $G \subseteq A$ si definiscono due operatori fondamentali.

La *conoscenza comune* C_G φ ("è di conoscenza comune nel gruppo G che φ") è definita semanticamente come il punto fisso dell'operazione "tutti i membri di G sanno che φ, e tutti sanno che tutti sanno che φ, e così via all'infinito". Formalmente, si introduce la relazione

R_G^* := (∪_{i∈G} R_i)^*

dove (·)^* indica la chiusura transitiva e riflessiva, e si pone

M, w ⊨ C_G φ iff per ogni v tale che wR_G^* v, M, v ⊨ φ

La conoscenza comune non è riducibile a nessuna iterazione finita degli operatori K_i, e pertanto C_G estende genuinamente la capacità espressiva di ℒ_EL. Il suo assioma caratteristico, il cosiddetto *assioma di punto fisso*, è

(FP) C_G φ ↔ E_G(φ ∧ C_G φ)

dove E_G φ := ∧_{i∈G} K_i φ ("tutti i membri di G sanno φ"). L'assioma FP afferma che la conoscenza comune è un punto fisso dell'operatore "φ e tutti lo sanno": C_G φ se e soltanto se tutti sanno (φ e la conoscenza comune di φ). Si aggiunge la regola di induzione:

(Ind) Se ⊢ ψ → E_G(φ ∧ ψ) allora ⊢ ψ → C_G φ

La *conoscenza distribuita* D_G φ ("se i membri del gruppo G mettessero insieme tutta la loro conoscenza individuale, saprebbero φ") è definita tramite la relazione

R_G^∩ := ∩_{i∈G} R_i

e la clausola semantica

M, w ⊨ D_G φ iff per ogni v tale che wR_G^∩ v, M, v ⊨ φ

La conoscenza distribuita è meno restrittiva della conoscenza comune: D_G φ può sussistere anche quando nessun singolo membro di G conosce φ, purché la conoscenza di φ sia "distribuita" tra i vari agenti in modo tale da essere logicamente recuperabile dalla loro unione. È un operatore particolarmente rilevante per la crittografia e per la teoria dei protocolli, dove ciò che il gruppo "sa" collettivamente è superiore alla somma delle conoscenze individuali.

I.6 Bisimulazione e invarianza

Un concetto tecnico di grande rilievo nella logica epistemica, che tornerà con forza nella DEL, è quello di *bisimulazione*. Una bisimulazione tra due modelli M = (W, {R_i}, V) e M' = (W', {R_i'}, V') è una relazione Z ⊆ W × W' tale che, per ogni (w, w') ∈ Z:

(atom) per ogni p ∈ Prop : w ∈ V(p) iff w' ∈ V'(p)
(forth) per ogni i ∈ A, per ogni v ∈ W con wR_i v,
esiste v' ∈ W' con w'R_i' v' e (v, v') ∈ Z
(back) per ogni i ∈ A, per ogni v' ∈ W' con w'R_i' v',
esiste v ∈ W con wR_i v e (v, v') ∈ Z

Due coppie mondo-modello (M, w) e (M', w') sono bisimili, scritto M, w ↔ M', w', se esiste una bisimulazione Z con (w, w') ∈ Z.

Teorema I.2 (Invarianza per bisimulazione): Se M, w ↔ M', w', allora per ogni φ ∈ ℒ_EL:

M, w ⊨ φ iff M', w' ⊨ φ

La dimostrazione procede per induzione sulla struttura di φ. Il caso base per le variabili proposizionali è immediato dalla clausola (atom). Il caso dell'operatore di conoscenza K_i usa le clausole (forth) e (back): se M, w ⊨ K_i φ e w'R_i' v', allora per (back) esiste v con wR_i v e (v, v') ∈ Z; per ipotesi M, v ⊨ φ, e per ipotesi induttiva M', v' ⊨ φ. L'altra direzione è analoga.

Questo teorema ha una duplice importanza. Semanticamente, dice che la logica epistemica è incapace di distinguere tra strutture bisimili: due modelli bisimili sono, dal punto di vista della logica, indistinguibili. Metodologicamente, fornisce uno strumento per dimostrare inesprimibilità: se una proprietà non è preservata dalla bisimulazione, essa non è esprimibile in ℒ_EL.

I.7 Limiti della logica epistemica statica e la necessità della dinamica

La logica epistemica classica, nella sua versione S5 multi-agente arricchita da conoscenza comune e distribuita, descrive con grande precisione la struttura *sincronica* della conoscenza: in un dato stato del mondo, chi sa cosa e in che relazione stanno le conoscenze dei diversi agenti. Ciò che rimane radicalmente fuori dalla portata di questo strumento è la dimensione *diacronica* del processo epistemico: come cambia lo stato epistemico degli agenti quando avviene qualcosa? Quando ricevono un messaggio, osservano un evento, o giungono a sapere che qualcun altro è venuto a conoscenza di qualcosa?

I modelli di Kripke statici non hanno le risorse per rispondere a queste domande. Essi descrivono situazioni epistemiche ma non trasformazioni tra situazioni: sono fotografie, non film. La necessità di andare oltre la statica, di dotare la logica

epistemica di operatori dinamici che catturino il cambiamento epistemico indotto dalla comunicazione e dall'azione, è il motore teorico che alimenta l'intero programma della DEL.

Il primo e più naturale tipo di evento epistemico da formalizzare è l'annuncio pubblico: un'entità esterna (o uno degli agenti) rende nota a tutti gli agenti una proposizione φ che è vera. Dopo questo annuncio, lo stato epistemico cambia: tutti sanno ora che φ, e questo fatto è a sua volta di conoscenza comune. Formalizzare questo cambiamento con precisione è il contributo fondamentale di Plaza, a cui il secondo capitolo è interamente dedicato.

Capitolo II — La Logica degli Annunci Pubblici: Plaza e l'Origine della DEL

II.1 Il contributo di Plaza e la sua collocazione storica

Nel 1989, in un articolo presentato al quarto simposio internazionale sulle metodologie per i sistemi intelligenti (Methodologies for Intelligent Systems), Jan Plaza pubblica un breve ma densissimo contributo intitolato *Logics of Public Communications*. Questo scritto, rimasto per qualche anno ai margini dell'attenzione specialistica prima di essere riconosciuto come uno dei testi fondativi della DEL, compie una mossa teorica di notevole audacia: introduce nel linguaggio della logica epistemica un operatore dinamico esplicito, l'operatore di *annuncio pubblico*, e ne fornisce una semantica rigorosa e un sistema assiomatico completo.

L'intuizione di fondo è semplice quanto feconda. Si immagini che in una situazione epistemica descritta da un modello M un agente — o un'entità esterna — renda pubblicamente nota a tutti gli agenti una proposizione φ che è attualmente vera. L'effetto di questo annuncio è duplice: da un lato, tutti gli agenti vengono a sapere che φ; dall'altro, diventano consapevoli che tutti gli altri sanno che φ, e che tutti sanno che tutti sanno, e così via — ovvero, φ diventa di conoscenza comune. Plaza formalizza questo processo come un'operazione di *restrizione del modello*: l'annuncio pubblico di φ elimina dal modello tutti i mondi in cui φ è falsa, poiché dopo l'annuncio nessun agente razionale può più ritenere possibile di trovarsi in un mondo in cui φ non vale.

La semplicità di questa idea semantica è ingannevole: essa introduce di fatto un'operazione di trasformazione dei modelli che costituisce il prototipo di una vasta famiglia di aggiornamenti epistemici che la DEL avrebbe esplorato nei decenni successivi.

II.2 Il linguaggio della logica degli annunci pubblici

Il linguaggio $\mathcal{L}$_PAL della *Public Announcement Logic* estende $\mathcal{L}$_EL con un nuovo tipo di formula atomica dinamica. La grammatica è

$$\varphi ::= p \mid \bot \mid \neg\varphi \mid (\varphi \wedge \psi) \mid K_i \varphi \mid [\varphi!]\psi$$

dove $p \in Prop$ e $i \in A$. La formula $[\varphi!]\psi$ si legge: "dopo l'annuncio pubblico di φ, ψ è vera" oppure "necessariamente, se φ viene annunciato pubblicamente, allora ψ vale successivamente". Il simbolo ! denota il carattere *assertivo* e *pubblico* dell'aggiornamento: si tratta di un annuncio che tutti sentono, che tutti sanno che tutti sentono, e così via.

L'operatore duale è definito come

$$\langle\varphi!\rangle\psi \;:=\; \neg[\varphi!]\neg\psi$$

e si legge: "è possibile che φ venga annunciato pubblicamente e che ψ valga successivamente", o equivalentemente "φ è annunciabile e dopo il suo annuncio ψ è vera". La dualità tra $[\varphi!]$ e $\langle\varphi!\rangle$ rispecchia la dualità standard tra necessità e possibilità nella logica modale.

Occorre notare subito che φ in $[\varphi!]\psi$ è essa stessa una formula di $\mathcal{L}$_PAL, non necessariamente priva di operatori dinamici: questo conferisce al linguaggio una notevole ricchezza espressiva, consentendo di annidare annunci all'interno di altri annunci. La profondità di annuncio di una formula — il numero massimo di annunci annidati — è definita per induzione in maniera analoga alla profondità modale.

II.3 La semantica relazionale per PAL

La semantica di $\mathcal{L}$_PAL è definita a partire dalla semantica epistemica standard di Kripke con l'aggiunta della clausola per l'operatore di annuncio. Si lavora con la stessa nozione di modello $M = (W, \{R_i\}_{i\in A}, V)$ introdotta nel capitolo precedente, con le R_i relazioni di equivalenza per ottenere la logica S5.

L'operazione semantica centrale è la *restrizione* del modello a una formula. Dato un modello M e una formula $\varphi \in \mathcal{L}$_PAL, il *modello ristretto* $M|\varphi$ è definito come

$$M|\varphi = (W^\varphi, \{R_i^\varphi\}_{i\in A}, V^\varphi)$$

dove

$$W^\varphi = \{ v \in W : M, v \vDash \varphi \}$$
$$R_i^\varphi = R_i \cap (W^\varphi \times W^\varphi)$$
$$V^\varphi(p) = V(p) \cap W^\varphi$$

In parole: W^φ è l'insieme dei mondi del modello originale in cui φ è vera; le relazioni di accessibilità vengono ristrette a questo sottoinsieme; la valutazione viene

analogamente ristretta. Si richiede, per buona definizione della semantica, che W^φ sia non vuoto, il che equivale a richiedere che φ sia vera in almeno un mondo del modello — ovvero che φ sia *annunciabile* in M.

La relazione di soddisfazione estesa a ℒ_PAL aggiunge alla definizione del capitolo precedente la clausola:

M, w ⊨ [φ!]ψ iff M, w ⊨ φ implica M|φ, w ⊨ ψ

Questa clausola merita una lettura attenta. Il connettivo materiale "implica" è essenziale: se φ è falsa in w, allora [φ!]ψ è vacuamente vera in w, qualunque sia ψ. La ragione è che se φ è falsa, l'annuncio di φ non può avvenire in w (un annuncio pubblico annuncia sempre qualcosa di vero), e dunque la formula è verificata per vacuità. Quando invece φ è vera in w, l'annuncio è possibile e ha luogo effettivamente: si elimina dal modello ogni mondo in cui φ manca, e si verifica se ψ vale nel mondo w del modello ridotto.

L'operatore duale ha la clausola

M, w ⊨ ⟨φ!⟩ψ iff M, w ⊨ φ e M|φ, w ⊨ ψ

che stabilisce: l'annuncio di φ è *possibile* in w (φ è vera) e, una volta eseguito, ψ è vera.

II.4 Gli assiomi di riduzione e la loro dimostrazione

Il risultato tecnico più importante della PAL è il *teorema di riduzione*, che Plaza stabilisce e van Ditmarsch, van der Hoek e Kooi sistematizzano nella loro trattazione. Esso afferma che ogni formula di ℒ_PAL è equivalente a una formula di ℒ_EL, ovvero che l'aggiunta degli operatori dinamici non aumenta la potenza espressiva della logica (anche se fornisce una notazione molto più compatta per certe proprietà).

Il teorema è dimostrato mediante un insieme di *assiomi di riduzione* (o schemi di riduzione), i quali forniscono un algoritmo di eliminazione degli operatori [φ!] dalla struttura sintattica delle formule. Si enumerano qui di seguito, accompagnati da una verifica semantica dettagliata.

(R_at) Riduzione per variabili atomiche:

[φ!]p ↔ (φ → p)

Verifica: M, w ⊨ [φ!]p iff (M, w ⊨ φ implica M|φ, w ⊨ p) iff (M, w ⊨ φ implica w ∈ V^φ(p)) iff (M, w ⊨ φ implica w ∈ V(p)) iff (M, w ⊨ φ implica M, w ⊨ p) iff M, w ⊨ φ → p. Il passaggio cruciale è che V^φ(p) = V(p) ∩ W^φ, e poiché w ∈ W^φ iff M, w ⊨ φ, si ha w ∈ V^φ(p) iff w ∈ V(p) (quando w ∈ W^φ). ▮

(R_⊥) Riduzione per la costante falsa:

$[\varphi!]\bot \leftrightarrow \neg\varphi$

Questo è un caso particolare di (R_at): l'annuncio di φ porta alla falsità se e soltanto se φ stessa era falsa (e dunque l'annuncio non era possibile).

(R_¬) Riduzione per la negazione:

$[\varphi!]\neg\psi \leftrightarrow (\varphi \rightarrow \neg[\varphi!]\psi)$

Verifica: M, w ⊨ [φ!]¬ψ iff (M, w ⊨ φ implica M|φ, w ⊨ ¬ψ) iff (M, w ⊨ φ implica M|φ, w ⊭ ψ) iff (M, w ⊨ φ implica ¬(M|φ, w ⊨ ψ)) iff (M, w ⊨ φ implica ¬(M, w ⊨ [φ!]ψ quando M, w ⊨ φ)). Ora se M, w ⊨ φ, allora M, w ⊨ [φ!]ψ iff M|φ, w ⊨ ψ. Dunque l'intera condizione equivale a: M, w ⊨ φ implica M, w ⊨ ¬[φ!]ψ, ovvero M, w ⊨ φ → ¬[φ!]ψ. ▌

(R_∧) Riduzione per la congiunzione:

$[\varphi!](\psi \wedge \chi) \leftrightarrow ([\varphi!]\psi \wedge [\varphi!]\chi)$

Verifica: Immediata dalla distributività dell'implicazione materiale (A → (B ∧ C)) ↔ ((A → B) ∧ (A → C)) e dalla clausola semantica di [φ!]. ▌

(R_K) Riduzione per l'operatore di conoscenza:

$[\varphi!]K_i\,\psi \leftrightarrow (\varphi \rightarrow K_i(\varphi \rightarrow [\varphi!]\psi))$

Questo è lo schema di riduzione più sostanzioso e filosoficamente più significativo. Vale la pena di verificarlo con attenzione.

Verifica: Si ha

M, w ⊨ [φ!]K_i ψ
iff M, w ⊨ φ implica M|φ, w ⊨ K_i ψ
iff M, w ⊨ φ implica (per ogni v ∈ W^φ con wR_i^φ v, M|φ, v ⊨ ψ)
iff M, w ⊨ φ implica (per ogni v ∈ W^φ con wR_i v, M|φ, v ⊨ ψ)

dove si usa che R_i^φ = R_i ∩ (W^φ × W^φ), quindi wR_i^φ v iff wR_i v e v ∈ W^φ. Continuando:

iff M, w ⊨ φ implica (per ogni v con wR_i v: M, v ⊨ φ implica M|φ, v ⊨ ψ)
iff M, w ⊨ φ implica (per ogni v con wR_i v: M, v ⊨ φ → [φ!]ψ)

dove nell'ultimo passaggio si usa che M, v ⊨ φ implica (M|φ, v ⊨ ψ iff M, v ⊨ [φ!]ψ), e se M, v ⊭ φ allora φ → [φ!]ψ è vacuamente vera in v. Dunque l'intera condizione equivale a

M, w ⊨ φ implica M, w ⊨ K_i(φ → [φ!]ψ)
iff M, w ⊨ φ → K_i(φ → [φ!]ψ) ▮

L'assioma (R_K) ha una lettura epistemica precisa: dopo l'annuncio pubblico di φ, l'agente i sa ψ se e soltanto se (a) φ era vera e (b) prima dell'annuncio i già sapeva che φ implica ψ. Condizione (b) si legge: in ogni mondo che i riteneva possibile, se φ vale allora ψ varrebbe dopo l'annuncio di φ. Questo esprime il fatto che l'annuncio informa gli agenti di φ, e che quanto sanno dopo dipende da quanto sapevano prima riguardo alle conseguenze di φ.

II.5 Il teorema di riduzione

Gli assiomi di riduzione forniscono un algoritmo di traduzione: data una formula φ ∈ ℒ_PAL con profondità di annuncio n > 0, si applica uno degli schemi (R_at), (R_¬), (R_⋀), (R_K) per abbassare la profondità di annuncio di almeno un'occorrenza dell'operatore [·!], ottenendo una formula equivalente. Iterando il processo fino a esaurimento di tutti gli operatori dinamici, si ottiene una formula di ℒ_EL equivalente all'originale.

Teorema II.1 (Riduzione, Plaza – van Ditmarsch–van der Hoek–Kooi): Per ogni φ ∈ ℒ_PAL esiste una formula φ^° ∈ ℒ_EL tale che ⊨ φ ↔ φ^°. La traduzione φ ↦ φ^° è effettiva (calcolabile in tempo finito) e la sua complessità è al più esponenziale nella profondità di annuncio di φ.

Dimostrazione: Per induzione sulla profondità di annuncio dp(φ). Se dp(φ) = 0, si pone φ^° = φ. Se dp(φ) > 0, si applica l'assioma di riduzione appropriato alla struttura principale di φ: ogni schema sostituisce φ con una formula equivalente avente profondità di annuncio strettamente minore (si noti che (R_K) introduce [φ!]ψ con la stessa formula d'annuncio φ ma applicata a ψ, formula che ha profondità di annuncio inferiore a K_i ψ all'interno di [φ!]K_i ψ). Per ipotesi induttiva, si conclude la riduzione. La crescita esponenziale deriva dal fatto che (R_¬) e (R_K) duplicano la formula d'annuncio φ nell'espressione risultante. ▮

Una conseguenza immediata è che PAL è *semanticamente conservativa* su ℒ_EL: aggiungere gli operatori di annuncio non esprime nessuna proprietà che non fosse già esprimibile nel frammento statico. Questa equivalenza espressiva è stata inizialmente vista come una limitazione, ma va intesa piuttosto come un risultato di coerenza: la DEL aggiunge *compattezza* (certe proprietà si esprimono molto più agevolmente con gli operatori dinamici) senza alterare la struttura espressiva di fondo.

II.6 Il sistema assiomatico e la completezza

Il sistema assiomatico per PAL su modelli S5 consiste di:

(Ax_EL) tutti gli assiomi e le regole di S5 (per tutti gli agenti $i \in A$)
(R_at) $[\varphi!]p \leftrightarrow (\varphi \rightarrow p)$
(R_¬) $[\varphi!]\neg\psi \leftrightarrow (\varphi \rightarrow \neg[\varphi!]\psi)$
(R_∧) $[\varphi!](\psi \wedge \chi) \leftrightarrow ([\varphi!]\psi \wedge [\varphi!]\chi)$
(R_K) $[\varphi!]K_i\,\psi \leftrightarrow (\varphi \rightarrow K_i(\varphi \rightarrow [\varphi!]\psi))$
(Nec_!) Se $\vdash \varphi$ allora $\vdash [\psi!]\varphi$ (necessitazione per gli annunci)

Teorema II.2 (Correttezza e completezza di PAL su S5): Il sistema assiomatico sopra descritto è suono e completo rispetto alla classe dei modelli di Kripke S5 con operatori di annuncio interpretati come sopra.

Strategia della dimostrazione: La correttezza si verifica schema per schema: si è già dimostrato che tutti gli assiomi di riduzione sono tautologie valide su ogni modello S5. La completezza si ricava dalla completezza di S5 mediante il teorema di riduzione: se $\vDash \varphi$, allora $\vDash \varphi^\circ$ (poiché φ e φ° sono equivalenti), e quindi $\vdash_{\{S5\}} \varphi^\circ$ *(per completezza di S5); infine, poiché* $\varphi \leftrightarrow \varphi^\circ$ *è deducibile nel sistema PAL per applicazione degli assiomi di riduzione, si ottiene* $\vdash_{\{PAL\}} \varphi$. ∎

Questa strategia di dimostrazione — dedurre la completezza di un sistema dinamico dalla completezza del sistema statico sottostante mediante i teoremi di riduzione — diventerà un paradigma metodologico ricorrente nell'intera tradizione della DEL.

II.7 Il paradosso di Moore e l'annunciabilità

Un aspetto sottile della semantica degli annunci pubblici emerge dall'analisi delle proposizioni *mooriane*. Una proposizione di tipo Moore è una formula del tipo

$$\varphi_M := p \wedge \neg K_i\, p$$

che afferma "p è vera e l'agente i non lo sa". In molti modelli epistemici questa formula è vera: si pensi a una situazione in cui p è un fatto del mondo ignoto all'agente. Tuttavia, se si tenta di annunciare φ_M pubblicamente, si ottiene qualcosa di paradossale: il solo atto dell'annuncio rende φ_M falsa, perché dopo l'annuncio l'agente i sa che p (poiché sa che φ_M, e quindi sa che p è vero). In simboli:

$\vDash \langle \varphi_M! \rangle \top$ (l'annuncio è possibile, cioè φ_M è annunciabile)

ma

$\vDash \neg\langle \varphi_M! \rangle K_i \bot$ (e dopo l'annuncio, l'agente non cade in contraddizione)

e soprattutto

$\vDash \langle \varphi_M! \rangle \neg K_i \bot$

Ciò che si può dimostrare è più sottile: in un modello M in cui M, w ⊨ φ_M, dopo l'annuncio di φ_M nel mondo w si ottiene M|φ_M, w ⊨ ¬φ_M. Ovvero φ_M non sopravvive al proprio annuncio: è vera prima ma falsa dopo. Calcolando esplicitamente con (R_K):

$$[\varphi_M!]K_i\, p \;\leftrightarrow\; \varphi_M \rightarrow K_i(\varphi_M \rightarrow [\varphi_M!]p)$$
$$\leftrightarrow\; \varphi_M \rightarrow K_i(\varphi_M \rightarrow p) \qquad [\text{da (R_at): } [\varphi_M!]p \leftrightarrow \varphi_M \rightarrow p]$$

Ora φ_M → p è equivalente a (p ∧ ¬K_i p) → p, che è una tautologia. Dunque

$$[\varphi_M!]K_i\, p \;\leftrightarrow\; \varphi_M \rightarrow K_i\top \;\leftrightarrow\; \varphi_M \rightarrow \top \;\leftrightarrow\; \top$$

Pertanto ⊨ [φ_M!]K_i p: dopo l'annuncio di φ_M, l'agente i sa che p. Ma φ_M = p ∧ ¬K_i p, il che implica che dopo l'annuncio ¬K_i p è diventata falsa, e dunque φ_M stessa è diventata falsa. Le proposizioni mooriane sono *annunciabili ma non persistenti*: il loro annuncio le rende false. Questo fenomeno non è un difetto della teoria ma un risultato corretto e rilevante, che cattura l'asimmetria tra verità e conoscenza.

II.8 Sequenze di annunci e composizione

La logica PAL permette naturalmente di ragionare su *sequenze* di annunci pubblici. Una sequenza di annunci φ_1, φ_2, …, φ_n si formalizza come annidamento di operatori dinamici:

$$[\varphi_1!][\varphi_2!]\ldots[\varphi_n!]\psi$$

La semantica interpreta questa formula come: "dopo aver annunciato φ_1, poi φ_2 nel modello risultante, poi φ_3, e così via fino a φ_n, la formula ψ vale". Grazie agli assiomi di riduzione, qualunque sequenza di annunci può essere ridotta a un singolo annuncio composto: esiste una formula $\chi \in \mathcal{L}_PAL$ tale che

$$\vDash\; [\varphi_1!][\varphi_2!]\ldots[\varphi_n!]\psi \;\leftrightarrow\; [\chi!]\psi$$

La composizione si ricava per applicazione iterata di (R_¬), (R_∧), (R_K) e delle formule di annuncio. In particolare, per la composizione di due annunci:

$$[\varphi!][\psi!]\chi \;\leftrightarrow\; [\varphi \wedge [\varphi!]\psi\,!]\chi$$

dove φ ∧ [φ!]ψ si legge: "φ vale e, dopo l'annuncio di φ, ψ vale". Questo unico annuncio composto ha lo stesso effetto degli annunci successivi φ e ψ. Si verifica con la semantica: i mondi che sopravvivono all'annuncio φ ∧ [φ!]ψ sono quelli in cui vale φ

e in cui vale [φ!]ψ, ovvero quelli w ∈ W^φ tali che M|φ, w ⊨ ψ — che è esattamente la condizione per sopravvivere alla restrizione doppia M|φ|ψ.

II.9 PAL con conoscenza comune

L'estensione di PAL con l'operatore di conoscenza comune C_G pone una difficoltà tecnica non banale: l'assioma di riduzione per [φ!]C_G ψ non si ottiene semplicemente applicando la strategia usata per K_i. Il risultato è uno schema che coinvolge la conoscenza comune stessa e non è puramente eliminatorio nel senso di ridurre la profondità di annuncio in un solo passo:

[φ!]C_G ψ ↔ φ → C_G(φ → [φ!]ψ)

Verifica: Si ha M, w ⊨ [φ!]C_G ψ iff M, w ⊨ φ implica M|φ, w ⊨ C_G ψ. Ora M|φ, w ⊨ C_G ψ iff per ogni v con wR_G^{,φ} *v (chiusura transitiva riflessiva di* ∪_{*i*∈*G*} *R_i^φ), si ha M|φ, v ⊨ ψ. La chiusura R_G^{,φ}* è la restrizione a W^φ della chiusura R_G^, *poiché R_i^φ = R_i ↾ W^φ. Dunque per ogni v raggiungibile da w tramite R_G^* passando solo per mondi in W^φ, si ha M|φ, v ⊨ ψ. Questo equivale a: per ogni v con wR_G^* v tale che ogni passo intermedio è in W^φ, M, v ⊨ φ e M, v ⊨ [φ!]ψ. Si può riformulare come: M, w ⊨ φ implica M, w ⊨ C_G(φ → [φ!]ψ), ovvero M, w ⊨ φ → C_G(φ → [φ!]ψ). ∎

Si noti che a destra dello schema compare ancora [φ!]ψ — l'operatore dinamico non è eliminato in un solo passo, ma il suo argomento ha profondità modale ridotta. Applicando il teorema di riduzione per C_G (che si ottiene da quello per K_i per iterazione transfinita), si conclude che anche ℒ_PAL+C_G è riducibile a ℒ_EL+C_G, preservando il risultato di equivalenza espressiva.

II.10 Limiti della PAL e la strada verso Gerbrandy

La Public Announcement Logic è un risultato teorico di grande eleganza, ma ha limiti intrinseci che la tradizione successiva avrebbe avvertito con crescente urgenza. Il limite principale è che PAL modella esclusivamente annunci *pubblici* nel senso più forte: tutti gli agenti ricevono la stessa informazione, tutti sanno di averla ricevuta, tutti sanno che tutti lo sanno. In realtà, la comunicazione epistemica è quasi sempre più selettiva: si inviano messaggi privati, si fanno segnali che solo alcuni osservano, si effettuano azioni il cui effetto informativo è diverso per diversi agenti.

Il passaggio da annunci pubblici a comunicazioni private — e più in generale a *eventi epistemici* di qualunque tipo — richiede una formalizzazione più sofisticata della struttura degli aggiornamenti. È questa la direzione in cui si muovono, con strategie diverse ma convergenti, Jelle Gerbrandy e successivamente Alexandru Baltag, Lawrence Moss e Slawomir Solecki, prima che la sintesi di van Ditmarsch, van der Hoek e Kooi consolidasse l'intero edificio teorico in una forma matura e sistematica.

La domanda che orienta questa successiva ricerca può essere formulata così: se nella PAL l'aggiornamento epistemico è modellato da un'operazione di restrizione del

modello (si eliminano i mondi in cui φ è falsa), quale struttura formale è necessaria per modellare aggiornamenti in cui agenti diversi ricevono informazioni diverse, o in cui l'azione stessa è conoscuta in modo differente dai diversi osservatori? La risposta, come si vedrà nel capitolo successivo, richiede di *aggiungere struttura agli eventi* — di trattare gli aggiornamenti non come semplici formule ma come oggetti Kripkiani a loro volta dotati di mondi, relazioni e precondizioni.

Capitolo III — Da Gerbrandy ai Modelli d'Azione: La Sintesi di van Ditmarsch, van der Hoek e Kooi

III.1 Il problema della comunicazione parziale e la svolta di Gerbrandy

Jelle Gerbrandy, nella sua dissertazione dottorale difesa all'Università di Amsterdam nel 1999 (*Bisimulations on Planet Kripke*), prende le mosse da un'osservazione che, enunciata nella sua forma più nuda, suona quasi ovvia: nella comunicazione reale, non tutti gli agenti vengono informati delle stesse cose, e non tutti sanno con certezza quali informazioni altri agenti abbiano ricevuto. Un agente che sussurra qualcosa a un secondo in presenza di un terzo mette in atto una comunicazione asimmetrica: il secondo riceve l'informazione, il primo sa di averla trasmessa, il terzo osserva che qualcosa è avvenuto ma non ne conosce il contenuto. Ciascuno dei tre si trova in uno stato epistemico diverso dopo l'evento, e le rispettive concezioni degli stati epistemici degli altri sono anch'esse differenti.

La PAL di Plaza è del tutto inadeguata a catturare questa ricchezza. Nell'annuncio pubblico l'evento comunicativo è omogeneo e trasparente: tutti ricevono la medesima proposizione φ, tutti sanno che tutti l'hanno ricevuta, e l'aggiornamento consiste nella semplice eliminazione dei mondi in cui φ è falsa. Non vi è spazio in questo schema per la comunicazione privata, per gli eventi parzialmente osservabili, o per situazioni in cui agenti diversi interpretano diversamente ciò che hanno osservato.

Gerbrandy propone di affrontare questo problema estendendo la struttura degli eventi. Mentre nella PAL un evento comunicativo è semplicemente una formula φ (la proposizione annunciata), in Gerbrandy un evento è anch'esso un *oggetto Kripkiano*: un modello dotato di mondi (che rappresentano diverse possibilità per l'evento stesso), relazioni di accessibilità (che codificano l'incertezza degli agenti riguardo a quale evento sia effettivamente occorso), e precondizioni (che specificano in quali mondi del modello fattuale l'evento è applicabile). Questa idea — l'idea che gli aggiornamenti epistemici debbano essere modellati con la stessa categoria di struttura usata per modellare gli stati epistemici — è il nucleo conceptuale più profondo dell'approccio di Gerbrandy, e verrà radicalizzata da Baltag, Moss e Solecki nella proposta dei modelli d'azione.

III.2 I modelli d'azione di Baltag, Moss e Solecki

Alexandru Baltag, Lawrence Moss e Slawomir Solecki, in un articolo del 1998 presentato al convegno TARK (*The Logic of Public Announcements, Common Knowledge, and Private Suspicions*), introducono la struttura che diventerà il cuore tecnico della DEL matura: i *modelli d'azione* (action models), spesso abbreviati con la sigla BMS in onore dei tre autori.

Un *modello d'azione* su un insieme di agenti A è una struttura

$\mathfrak{A} = (\text{Act}, \{\approx_i\}_{i\in A}, \text{pre})$

dove Act è un insieme finito non vuoto di *eventi possibili* (o azioni), per ogni agente $i \in A$ la relazione $\approx_i \subseteq \text{Act} \times \text{Act}$ è la *relazione di indistinguibilità* dell'agente i su gli eventi (si assume che sia una relazione di equivalenza), e $\text{pre} : \text{Act} \to \mathcal{L}_{EL}$ è la funzione di *precondizione*, che assegna a ogni evento $\alpha \in \text{Act}$ una formula epistemica pre(α) che deve essere vera nel mondo del modello fattuale affinché l'evento α sia effettivamente eseguibile.

La struttura di $\mathfrak{A}$ va letta nel modo seguente. Un singolo evento concreto designato $\alpha\bigstar \in \text{Act}$ è l'evento che effettivamente occorre. Gli altri eventi in Act sono eventi *alternativi* che diversi agenti potrebbero, a seconda della loro posizione epistemica, credere di stare osservando. L'agente i non distingue tra due eventi α e β (ovvero $\alpha \approx_i \beta$) se non è in grado di stabilire, dalla sua prospettiva, se sia avvenuto α oppure β. La precondizione pre(α) specifica la condizione che il mondo fattuale deve soddisfare affinché α sia l'evento che effettivamente ha luogo.

Due esempi canonici illustrano il potere espressivo di questa struttura.

Esempio: annuncio pubblico come modello d'azione. Un annuncio pubblico della formula φ è rappresentato dal modello d'azione

$\mathfrak{A}_{\{\varphi!\}} = (\text{Act}, \{\approx_i\}, \text{pre})$

con $\text{Act} = \{\alpha\}$, l'unico evento possibile, $\approx_i = \{(\alpha,\alpha)\}$ per tutti gli agenti (nessuna incertezza sull'evento: tutti sanno quale evento è avvenuto), e $\text{pre}(\alpha) = \varphi$. Questo modello d'azione codifica esattamente la struttura dell'annuncio pubblico plaziano: un unico evento, noto a tutti, con precondizione φ. La PAL è un caso speciale della teoria BMS.

Esempio: comunicazione privata. Due agenti, 1 e 2, in presenza di un terzo agente 3. L'agente 1 sussurra qualcosa all'agente 2 che l'agente 3 può osservare come evento ma non decifrare. Si considera il caso in cui 1 potrebbe aver sussurrato "p è vera" oppure non aver detto nulla (o aver sussurrato qualcosa di irrilevante). Si ha

$\text{Act} = \{\alpha, \beta\}$

dove α è l'evento "1 sussurra a 2 che p è vera" e β è l'evento "nulla di sostanziale viene comunicato". Le relazioni di indistinguibilità sono

$\alpha \approx_1 \alpha$, $\beta \approx_1 \beta$ (1 sa quale dei due eventi è avvenuto)
$\alpha \approx_2 \alpha$, $\beta \approx_2 \beta$ (2 sa quale dei due eventi è avvenuto)
$\alpha \approx_3 \beta$, $\beta \approx_3 \alpha$ (3 non distingue i due eventi)

Le precondizioni sono $\mathrm{pre}(\alpha) = p$ e $\mathrm{pre}(\beta) = \top$. Questo modello d'azione rappresenta fedelmente la struttura epistemica della comunicazione privata: se α è l'evento effettivo, allora p deve valere nel mondo attuale; l'agente 3 non sa se α o β è avvenuto.

III.3 Il prodotto di aggiornamento

Il meccanismo formale mediante il quale un modello d'azione trasforma un modello epistemico è il *prodotto di aggiornamento* (update product), che costituisce il contributo tecnico più originale del framework BMS. Dato un modello epistemico $M = (W, \{R_i\}, V)$ e un modello d'azione $\mathfrak{A} = (\mathrm{Act}, \{\approx_i\}, \mathrm{pre})$, il prodotto di aggiornamento è un nuovo modello epistemico

$$M \otimes \mathfrak{A} = (W', \{R_i'\}, V')$$

definito come segue. L'insieme dei mondi del modello aggiornato è

$$W' = \{ (w, \alpha) \in W \times \mathrm{Act} : M, w \vDash \mathrm{pre}(\alpha) \}$$

ovvero le coppie (mondo del modello fattuale, evento possibile del modello d'azione) tali che la precondizione dell'evento è soddisfatta nel mondo fattuale. Si richiede che $(w, \alpha) \in W'$ affinché l'evento α sia effettivamente eseguibile in w.

Le relazioni di accessibilità nel modello aggiornato sono

$$(w, \alpha)\, R_i'\, (v, \beta) \quad \text{iff} \quad w R_i v \ \text{ e } \ \alpha \approx_i \beta$$

Due coppie (w, α) e (v, β) sono accessibili per l'agente i se e soltanto se i due mondi fattuali w e v erano R_i-accessibili nel modello originale e i due eventi α e β erano indistinguibili per l'agente i nel modello d'azione. Questa clausola cattura l'interazione tra incertezza sul mondo e incertezza sull'evento: dopo l'aggiornamento, l'agente i non può distinguere (w, α) da (v, β) se prima non poteva distinguere w da v riguardo al mondo fattuale e non poteva distinguere α da β riguardo all'evento.

La funzione di valutazione del modello aggiornato è semplicemente

$$V'(p) = \{ (w, \alpha) \in W' : w \in V(p) \}$$

Le variabili proposizionali mantengono il loro valore nel mondo fattuale, indipendentemente dall'evento.

Definizione III.1: Data una coppia (M, w) con w ∈ W e un modello d'azione 𝔄 con evento designato α★ ∈ Act, se M, w ⊨ pre(α★) allora l'*aggiornamento* di (M, w) con (𝔄, α★) è la coppia (M ⊗ 𝔄, (w, α★)).

Il mondo reale nell'aggiornamento è dunque (w, α★): il mondo fattuale è w e l'evento effettivamente occorso è α★.

III.4 Il linguaggio e la semantica della DEL completa

Il linguaggio ℒ_DEL, che estende ℒ_EL con operatori dinamici per i modelli d'azione, è definito dalla grammatica

φ ::= p | ⊥ | ¬φ | (φ ∧ ψ) | K_i φ | [(𝔄, α)]φ

dove (𝔄, α) è una coppia formata da un modello d'azione 𝔄 e da uno dei suoi eventi α. La formula [(𝔄, α)]φ si legge: "necessariamente, se l'evento α del modello d'azione 𝔄 occorre, allora φ vale dopo". La duale è ⟨(𝔄, α)⟩φ := ¬[(𝔄, α)]¬φ.

La clausola semantica è

M, w ⊨ [(𝔄, α)]φ iff M, w ⊨ pre(α) implica M ⊗ 𝔄, (w, α) ⊨ φ

Il parallelismo con la semantica di PAL è evidente: [φ!]ψ corrisponde a [(𝔄_{φ!}, α)]ψ con 𝔄_{φ!} come definito sopra. Ma la struttura BMS è enormemente più generale: l'evento può essere uno tra molti, gli agenti possono avere incertezze differenti su quale evento sia avvenuto, e le precondizioni possono essere formule epistemicamente ricche.

III.5 Gli assiomi di riduzione per i modelli d'azione

Anche per i modelli d'azione è possibile stabilire assiomi di riduzione che permettano di eliminare gli operatori dinamici [(𝔄, α)] a favore di formule del linguaggio statico ℒ_EL (arricchito eventualmente con conoscenza comune). Gli schemi sono del tutto analoghi a quelli della PAL, con la precondizione pre(α) al posto di φ e con una formula di conoscenza più articolata al posto di K_i.

(R_at^BMS):

[(𝔄, α)]p ↔ (pre(α) → p)

(R_¬^BMS):

$[(\mathfrak{A}, \alpha)]\neg\varphi \leftrightarrow (pre(\alpha) \rightarrow \neg[(\mathfrak{A}, \alpha)]\varphi)$

(R_⋀^BMS):

$[(\mathfrak{A}, \alpha)](\varphi \wedge \psi) \leftrightarrow ([(\mathfrak{A}, \alpha)]\varphi \wedge [(\mathfrak{A}, \alpha)]\psi)$

(R_K^BMS): Questo è lo schema più significativo. Per ogni evento $\alpha \in Act$:

$[(\mathfrak{A}, \alpha)]K_i\,\varphi \leftrightarrow (pre(\alpha) \rightarrow K_i \wedge_{\{\beta: \alpha \approx_i \beta\}} (pre(\beta) \rightarrow [(\mathfrak{A}, \beta)]\varphi))$

La lettura epistemica è la seguente: dopo l'esecuzione dell'evento α, l'agente i sa φ se e soltanto se, prima dell'evento, in ogni mondo che i riteneva possibile (e in ogni evento che i non distingueva da α), se l'evento alternativo β aveva la precondizione soddisfatta, allora φ avrebbe valso dopo quell'evento alternativo. In altri termini: i sa φ post-aggiornamento se i già sapeva che φ sarebbe valsa per tutti gli eventi alternativi epistemicamente compatibili con α.

Verifica di (R_K^BMS): Si ha

M, w ⊨ $[(\mathfrak{A}, \alpha)]K_i\,\varphi$
iff M, w ⊨ pre(α) implica $M \otimes \mathfrak{A}$, (w,α) ⊨ $K_i\,\varphi$
iff M, w ⊨ pre(α) implica $\forall$ (v,β) con (w,α)R_i'(v,β): $M \otimes \mathfrak{A}$, (v,β) ⊨ φ

Ora (w,α)R_i'(v,β) iff wR_i v e $\alpha \approx_i \beta$. Dunque la condizione diventa:

M, w ⊨ pre(α) implica $\forall$ v(wR_i v) $\forall$ β($\alpha \approx_i \beta$): (v,β) ∈ W' implica $M \otimes \mathfrak{A}$,(v,β) ⊨ φ

Poiché (v,β) ∈ W' iff M,v ⊨ pre(β), si riscrive come:

M, w ⊨ pre(α) implica $\forall$ v(wR_i v) $\forall$ β($\alpha \approx_i \beta$): M,v ⊨ pre(β) implica $M \otimes \mathfrak{A}$,(v,β) ⊨ φ
iff M, w ⊨ pre(α) implica $\forall$ v(wR_i v): M,v ⊨ $\wedge_{\{\beta:\alpha \approx_i \beta\}}$(pre(β) → $[(\mathfrak{A},\beta)]\varphi$)
iff M, w ⊨ pre(α) → $K_i(\wedge_{\{\beta:\alpha \approx_i \beta\}}(pre(\beta) \rightarrow [(\mathfrak{A},\beta)]\varphi))$ ▌

Questa verifica illumina la struttura profonda dell'aggiornamento BMS: la conoscenza post-aggiornamento dipende dalla conoscenza pre-aggiornamento non solo riguardo al mondo fattuale (come nella PAL) ma anche riguardo all'evento occorso, attraverso la quantificazione su tutti gli eventi alternativi β indistinguibili da α per l'agente i.

III.6 La monografia di van Ditmarsch, van der Hoek e Kooi

Hans van Ditmarsch, Wiebe van der Hoek e Barteld Kooi pubblicano nel 2007 il volume *Dynamic Epistemic Logic* presso Springer, che costituisce la prima trattazione sistematica e definitiva dell'intera tradizione. Il volume integra la PAL plaziana, i

modelli d'azione BMS, e una serie di estensioni importanti tra cui la logica dell'aggiornamento plausibile e la logica condizionale doxastica. Esso diventa immediatamente il testo di riferimento per il campo.

Il contributo sistematico di van Ditmarsch, van der Hoek e Kooi non si limita alla raccolta e all'organizzazione dei risultati precedenti: essi sviluppano con cura le questioni di decidibilità, completezza e complessità computazionale, trattano in dettaglio i protocolli di comunicazione e i giochi epistemici, e introducono una serie di raffinamenti tecnici che chiariscono i limiti espressivi e computazionali della DEL. Di seguito si espongono i contributi più originali di questa opera che non trovano antecedenti diretti nei lavori precedenti.

III.7 La logica della credenza e l'aggiornamento plausibile

Un'importante direzione di ricerca sistematizzata da van Ditmarsch, van der Hoek e Kooi riguarda il passaggio dalla *conoscenza* nel senso S5 alla *credenza* e all'aggiornamento epistemico in condizioni di incertezza più realistica. In molte situazioni praticamente rilevanti gli agenti non *sanno* le proposizioni in gioco ma le *credono* sulla base di una gerarchia di plausibilità: alcune possibilità sono ritenute più probabili di altre, e l'aggiornamento epistemico dovrebbe rispettare questa struttura di preferenza.

Questa idea è formalizzata mediante i *modelli di plausibilità* (plausibility models), noti anche come modelli di ordine lessicografico o modelli doxastici. Un modello di plausibilità è una struttura

$$M = (W, \{\leq_i\}_{i \in A}, V)$$

dove $\leq_i \subseteq W \times W$ è un *ordine di plausibilità* totale e transitivo per l'agente i — si legge $w \leq_i v$ come "l'agente i ritiene v almeno tanto plausibile quanto w". La relazione di credenza è definita a partire dall'ordine di plausibilità: l'agente i *crede* φ in w se φ è vera in tutti i mondi che i considera massimalmente plausibili tra quelli che ritiene possibili:

$$M, w \models B_i \varphi \quad \text{iff} \quad \forall v \in \mathrm{Min}_{\leq_i}(W_i(w)) : M, v \models \varphi$$

dove $W_i(w)$ è l'insieme dei mondi accessibili da w per l'agente i (la sua sfera doxastica) e $\mathrm{Min}_{\leq_i}(S)$ denota i minimi di S secondo $\leq_i$ (i mondi ritenuti più plausibili nell'insieme S). Questa semantica è quella dei *modelli di Spohn* nella versione sviluppata da Halpern, Shoham e Grove, e corrisponde assiomaticamente al sistema KD45.

L'aggiornamento epistemico sui modelli di plausibilità non è la semplice restrizione plaziana (eliminazione dei mondi in cui φ è falsa) ma una rielaborazione dell'ordine di plausibilità. Due regole di aggiornamento sono state studiate con particolare attenzione.

Aggiornamento condizionalizzante (Conditionalization): Dato l'annuncio di φ, la regola più semplice è eliminare i mondi in cui ¬φ e mantenere invariato l'ordine di plausibilità sui mondi rimanenti. Questa è la regola di Jeffrey:

$$\leq_i^{\varphi!} \;=\; \leq_i \restriction (W^{\varphi} \times W^{\varphi})$$

Corrisponde all'aggiornamento bayesiano condizionato quando le plausibilità sono interpretate probabilisticamente.

Aggiornamento lessicografico (Lexicographic Upgrade): Quando l'annuncio di φ è una *revisione* (il contenuto di φ contraddice le credenze correnti dell'agente), la semplice condizionalizzazione porta a risultati indesiderati: se l'agente credeva ¬φ, l'annuncio di φ richiederebbe di condizionalizzare su un insieme di mondi che l'agente riteneva tutti inferiori in plausibilità. L'aggiornamento lessicografico risolve questo problema riordinando tutti i mondi in cui φ è vera *sopra* i mondi in cui ¬φ vale, mantenendo l'ordine relativo interno ai due gruppi:

$$\begin{aligned} v \leq_i^{\Uparrow\varphi} u \quad \text{iff} \quad & (M,v \models \varphi \;\text{ e }\; M,u \models \varphi) \\ \vee \;\; & (M,v \models \varphi \;\text{ e }\; M,u \models \varphi \;\text{ e }\; v \leq_i u) \\ \vee \;\; & (M,v \not\models \varphi \;\text{ e }\; M,u \not\models \varphi \;\text{ e }\; v \leq_i u) \end{aligned}$$

Dopo un aggiornamento lessicografico con φ, i mondi in cui φ vale diventano quelli ritenuti più plausibili, seguiti da quelli in cui ¬φ vale. Questo corrisponde alla logica della *revisione* nel senso di AGM (Alchourrón–Gärdenfors–Makinson), e in particolare all'operazione di *revisione* piuttosto che di semplice *contrazione* o *espansione*.

Van Ditmarsch, van der Hoek e Kooi dimostrano che entrambi i tipi di aggiornamento sono catturabili nel framework BMS con opportune strutture di modelli d'azione su ordini di plausibilità, e forniscono sistemi assiomatici completi per i linguaggi risultanti.

III.8 La logica DEL e la teoria dei giochi

Un'altra linea di ricerca sviluppata sistematicamente da van Ditmarsch, van der Hoek e Kooi riguarda i *giochi epistemici* — situazioni in cui più agenti interagiscono strategicamente e la conoscenza di ciascuno delle mosse e degli stati del gioco è parziale o asimmetrica. La DEL fornisce gli strumenti per analizzare *announcements* come mosse di un gioco (giochi di informazione), *estrategie in equilibrio bayesiano* come stati di conoscenza reciproca, e la *struttura epistemica della razionalità comune* come punto fisso della conoscenza comune del fatto che tutti gli agenti sono razionali.

Il teorema di backward induction nei giochi finiti a informazione perfetta, in questa prospettiva, può essere visto come un risultato di DEL: è la conoscenza comune della razionalità degli agenti (opportunamente formalizzata nel linguaggio epistemico) che determina l'unico esito razionalmente ammissibile. La formalizzazione di questo argomento in $\mathcal{L}$_DEL è non banale e ha alimentato un lungo dibattito; van

Ditmarsch, van der Hoek e Kooi ne trattano le implicazioni in relazione alle proprietà del sistema assiomatico.

III.9 Decidibilità e complessità computazionale

La trattazione sistematica del volume del 2007 affronta con rigore le questioni di decidibilità e di complessità che la tradizione precedente aveva in parte trascurato. I risultati principali sono i seguenti.

Teorema III.1 (Decidibilità di PAL): Il problema di soddisfacibilità per $\mathcal{L}$_PAL (date una formula $\varphi \in \mathcal{L}$_PAL, esiste un modello S5 M e un mondo $w \in W$ tali che $M, w \models \varphi$?) è decidibile. La decidibilità segue direttamente dal teorema di riduzione: ogni formula di $\mathcal{L}$_PAL è equivalente a una formula di $\mathcal{L}$_EL, e la soddisfacibilità di $\mathcal{L}$_EL su modelli S5 è decidibile (in quanto S5 è un sistema modale decidibile, PSPACE-completo nel caso proposizionale multi-agente secondo il risultato di Ladner 1977 per il caso mono-agente e le sue generalizzazioni).

Teorema III.2 (Complessità di PAL): Il problema di soddisfacibilità per $\mathcal{L}$_PAL è PSPACE-completo. Il limite superiore deriva dal fatto che la traduzione in $\mathcal{L}$_EL può avere dimensione esponenziale, ma la soddisfacibilità di $\mathcal{L}$_EL è in PSPACE; il limite inferiore segue dalla PSPACE-hardness di $\mathcal{L}$_EL.

Teorema III.3 (Decidibilità di $\mathcal{L}$_DEL con modelli d'azione finiti): Il problema di soddisfacibilità per $\mathcal{L}$_DEL con modelli d'azione finiti è decidibile. Anche in questo caso la strategia consiste nel dimostrare l'esistenza di assiomi di riduzione e nel ricondurre la soddisfacibilità al sistema statico sottostante. La complessità è tuttavia superiore: la traduzione può introdurre una crescita non elementare nella dimensione della formula ridotta per modelli d'azione con struttura epistemica ricca.

Teorema III.4 (Completezza di $\mathcal{L}$_DEL): Il sistema assiomatico che estende S5 con gli schemi di riduzione (R_at^BMS), (R_¬^BMS), (R_$\wedge$^BMS), (R_K^BMS) e la regola di necessitazione $[(\mathfrak{A},\alpha)]$ è suono e completo rispetto alla classe dei modelli S5 con prodotto di aggiornamento BMS.

La dimostrazione di completezza usa la stessa strategia del caso PAL: la riduzione alle formule statiche e la completezza del sistema modale di partenza. La verifica che gli assiomi di riduzione sono derivabili nel sistema assiomatico esteso è il passo tecnico più delicato.

III.10 Bisimulazione per modelli d'azione e invarianza

Una questione concettualmente importante che van Ditmarsch, van der Hoek e Kooi affrontano è quella dell'invarianza per bisimulazione nel contesto dei modelli d'azione. Si definisce la bisimulazione per coppie (modello epistemico, modello d'azione) in modo analogo alla bisimulazione per modelli epistemici:

Una *bisimulazione per prodotti* tra $(M, \mathfrak{A})$ e $(M', \mathfrak{A}')$ è una relazione $Z \subseteq (W \times \mathrm{Act}) \times (W' \times \mathrm{Act}')$ tale che per ogni $((w,\alpha),(w',\alpha')) \in Z$:

(atom) $\forall p: w \in V(p)$ iff $w' \in V'(p)$
(pre) $M, w \vDash pre(\alpha)$ iff $M', w' \vDash pre'(\alpha')$
(forth) $\forall i: \forall (v,\beta)$ con $(w,\alpha)R_i'(v,\beta): \exists (v',\beta')$ con $(w',\alpha')(R_i')'(v',\beta')$ e $((v,\beta),(v',\beta')) \in Z$
(back) [analoga]

Teorema III.5 (Invarianza per bisimulazione di $\mathcal{L}$_DEL): Se $(M,w,\mathfrak{A},\alpha)$ e $(M',w',\mathfrak{A}',\alpha')$ sono in relazione di bisimulazione per prodotti, allora per ogni $\varphi \in \mathcal{L}$_DEL:

$M \otimes \mathfrak{A}, (w,\alpha) \vDash \varphi$ iff $M' \otimes \mathfrak{A}', (w',\alpha') \vDash \varphi$

Questo risultato ha implicazioni dirette per la teoria dell'inesprimibilità: proprietà delle strutture epistemiche dinamiche che non sono preservate dalla bisimulazione per prodotti non possono essere espresse in $\mathcal{L}$_DEL.

III.11 Limiti della DEL e prospettive aperte

La DEL nella sua formulazione BMS-vDHK, pur costituendo un risultato teorico di grande compiutezza, presenta limiti che la ricerca successiva ha messo progressivamente in luce. Il primo riguarda gli agenti *fallibili*: nella DEL standard si assume che le precondizioni degli eventi siano formule vere, ovvero che gli aggiornamenti siano basati su informazione corretta. Modellare aggiornamenti basati su informazione *falsa* (bugie, inganni, comunicazioni erronee) richiede un'estensione della struttura degli eventi.

Il secondo limite riguarda la *granularità temporale*: la DEL modella singoli passi di aggiornamento o composizioni finite di tali passi, ma non processi epistemici di durata indefinita come quelli che la teoria formale dell'apprendimento di Kelly si propone di analizzare. Nella DEL, ogni sequenza di aggiornamenti è finita e determinata in anticipo dalla struttura dei modelli d'azione. La teoria di Kelly affronta invece sequenze di osservazioni potenzialmente infinite e si chiede se un metodo d'indagine converga alla verità lungo tali sequenze. È questa differenza di prospettiva temporale — sincronica e finitistica nella DEL, diacronica e potenzialmente infinita in Kelly — che motiva la separazione delle due tradizioni e insieme la loro complementarità, come si chiarirà nel corso della seconda parte di questa monografia.

Capitolo IV — L'Indagine Affidabile: Da Gold a Kelly

IV.1 Il problema dell'induzione come problema matematico

L'epistemologia della scienza ha tradizionalmente posto il problema dell'induzione nella forma humiana: quali sono le condizioni in cui un'inferenza da casi osservati a generalizzazioni universali è razionalmente giustificata? La risposta popperiana, che sostituisce la verifica con la falsificazione come criterio di demarcazione e rigetta del tutto la nozione di giustificazione induttiva, ha dominato la metodologia scientifica per decenni. Essa tuttavia lascia insoddisfatte molte domande: se nessun metodo d'indagine è giustificato a pretendere la verità delle proprie congetture, in quale senso si può affermare che la scienza fa progresso? È possibile confrontare metodologie diverse secondo criteri di eccellenza che non presuppongano la verifica?

E.M. Gold, in un articolo del 1967 pubblicato su *Information and Control* e intitolato *Language Identification in the Limit*, apre una strada radicalmente diversa. Gold non cerca di *giustificare* l'induzione nel senso epistemologico tradizionale, ma si chiede se sia possibile *identificare* la risposta corretta *al limite* di una sequenza infinita di osservazioni. L'approccio è computazionale e matematico: si definisce con precisione che cosa significa che un metodo d'indagine *converge* alla risposta corretta, e si studia per quali classi di problemi e per quali metodi questa convergenza è garantita.

Kevin Kelly, in una serie di contributi che si estendono dal volume *The Logic of Reliable Inquiry* (1996) fino a numerosi articoli specialistici, prende l'approccio di Gold come punto di partenza e lo trasforma in qualcosa di molto più ampio e ambizioso: una teoria della razionalità epistemica nella sua dimensione diacronica, con implicazioni dirette per la metodologia scientifica, per la valutazione comparativa dei metodi d'indagine, e per la giustificazione formale di norme epistemiche come il rasoio di Occam.

IV.2 Il framework di Gold: identificazione al limite

Il contesto originale di Gold è l'identificazione di linguaggi formali. Si immagini un bambino che apprende la grammatica della propria lingua madre: riceve una sequenza (potenzialmente infinita) di esempi di frasi grammaticali (e, in alcune versioni del problema, di frasi non grammaticali), e deve congetturare quale sia la grammatica sottostante. Gold si chiede: esistono metodi di apprendimento che convergono alla grammatica corretta nel senso che, dopo un numero finito di esempi, formulano la congettura giusta e non la cambiano mai più?

Il framework astratto, nella generalizzazione che costituirà la base della teoria kellyiana, è il seguente.

Un *problema d'indagine* è specificato da:

(i) uno spazio di possibilità Ω (l'insieme delle ipotesi o delle teorie candidate)

(ii) una struttura di evidenza — una sequenza di osservazioni $e \in E^{\omega}$ (dove E è un insieme di possibili osservazioni e E^{ω} è l'insieme delle sequenze infinite)

(iii) una funzione di verità $V : \Omega \times E^{\omega} \to \{0, 1\}$ che, per ogni ipotesi $h \in \Omega$ e ogni sequenza di evidenza $\sigma \in E^{\omega}$, stabilisce se h è vera in σ

Un *metodo d'indagine* (o learner, o conjecturer) è una funzione

$$M : E^* \to \Omega \cup \{?\}$$

dove E^* è l'insieme di tutte le sequenze finite di osservazioni e ? rappresenta la risposta "nessuna congettura". Per ogni sequenza finita $e^n = (e_0, e_1, \ldots, e_{n-1}) \in E^*$, il metodo M produce una congettura $M(e^n) \in \Omega$ oppure si astiene (?).

Definizione IV.1 (Convergenza e identificazione al limite): Sia $\sigma = (e_0, e_1, e_2, \ldots) \in E^\omega$ una sequenza infinita di osservazioni. Si dice che il metodo M *converge* su σ se esiste $N \in \mathbb{N}$ tale che per tutti gli $n \geq N$, $M(e^n) = h$ per qualche ipotesi fissa $h \in \Omega$. Si dice che M *identifica al limite* l'ipotesi h su σ se converge su σ con valore limite h. Un metodo M *identifica una classe di ipotesi* $H \subseteq \Omega$ se, per ogni ipotesi $h \in H$ e per ogni sequenza σ compatibile con h (ovvero tale che $V(h, \sigma) = 1$), M identifica al limite h su σ.

Il risultato fondamentale di Gold riguarda la gerarchia di Chomsky: la classe dei linguaggi regolari e la classe dei linguaggi context-free (definiti positivamente, ovvero da sequenze di esempi grammaticali senza esempi negativi) non sono identificabili al limite, mentre lo sono per esempio le classi finite di linguaggi e alcune sottoclassi ben definite. Il teorema che esclude l'identificazione della classe dei linguaggi context-free è dimostrato per via diagonale: si costruisce una sequenza di evidenza che *enganna* qualunque metodo proposto, costringendolo a cambiare congettura infinitamente spesso o a convergere a un'ipotesi errata.

Teorema IV.1 (Gold 1967): La classe di tutti i linguaggi ricorsivamente enumerabili non è identificabile al limite da nessun metodo computazionalmente effettivo, nemmeno con esempi positivi e negativi.

La dimostrazione è una variante della diagonalizzazione cantorian-turingniana. Si numerino tutti i metodi computazionali $M_0, M_1, M_2, \ldots$ Si costruisce una sequenza di evidenza σ nel modo seguente: si inizia con un segmento finito σ_0 tale che $M_0(\sigma_0) = L_0$ per qualche linguaggio L_0. Si estende poi σ_0 a σ_1 cambiando il linguaggio "vero" in modo da forzare M_0 a sbagliare, poi si lascia che M_1 converga a qualcosa e si forza M_1 a sbagliare nel passo successivo, e così via. La costruzione produce una sequenza rispetto alla quale nessun metodo M_i converge alla risposta corretta.

IV.3 Il framework kellyiano: l'indagine affidabile

Kevin Kelly nella sua teoria sposta il baricentro della discussione dall'identificazione al limite (un criterio binario: converge o non converge) a una famiglia più ricca di criteri di performance epistemica. L'idea centrale è che l'eccellenza di un metodo d'indagine si misura non soltanto in termini di convergenza finale ma anche in termini di *efficienza*

della convergenza — quanti cambiamenti di congettura il metodo è costretto a effettuare lungo il cammino verso la risposta corretta?

Questa distinzione ha una motivazione filosofica profonda. Due metodi possono entrambi convergere alla risposta corretta, ma l'uno può cambiare congettura quattro volte mentre l'altro ne cambia mille: vi è chiaramente una differenza epistemica tra i due. Kelly formalizza questa intuizione mediante la nozione di *mind changes* (cambiamenti di mente) e ne fa la metrica fondamentale della teoria.

Definizione IV.2 (Cambiamenti di congettura): Sia M un metodo e $\sigma \in E^{\omega}$ una sequenza di evidenza. Il numero di *cambiamenti di congettura* di M su σ, denotato $mc(M, \sigma)$, è la cardinalità dell'insieme

$$\{ n \in \mathbb{N} : M(\sigma|n) \neq M(\sigma|(n+1)) \text{ e } M(\sigma|n) \neq ? \text{ e } M(\sigma|(n+1)) \neq ? \}$$

ovvero il numero di passi in cui M cambia la propria congettura da un valore non-? a un altro valore non-? (non si conta il passaggio dall'astensione a una prima congettura). Il numero $mc(M, \sigma)$ può essere un naturale finito o $+\infty$.

La *complessità epistemica* di un problema d'indagine misura il numero minimo di cambiamenti di congettura che qualunque metodo risolutivo deve essere disposto ad accettare nel caso peggiore:

Definizione IV.3 (Complessità in cambiamenti di congettura): La *complessità* del problema d'indagine $\langle \Omega, E^{\omega}, V, H \rangle$ rispetto alla classe di ipotesi H è

$$CC(H) = \inf_M \sup_{h \in H} \sup_{\sigma: V(h,\sigma)=1} mc(M, \sigma)$$

dove l'inf è preso su tutti i metodi che identificano H al limite. Se $CC(H) = k$, allora ogni metodo risolutivo deve effettuare almeno k cambiamenti di congettura su almeno qualche sequenza di evidenza, ma esiste un metodo che non ne effettua mai più di k.

IV.4 La topologia dello spazio delle ipotesi

Una delle innovazioni più originali di Kelly è l'adozione di un approccio *topologico* alla struttura dello spazio delle ipotesi. Questo passo — di notevole eleganza matematica — permette di connettere la teoria dell'indagine affidabile alla topologia generale e di derivare risultati di complessità mediante invarianti topologici.

Sia Ω uno spazio di ipotesi e sia E un insieme di osservazioni possibili. Si definisce la *topologia dell'evidenza* su Ω nel modo seguente. Per ogni sequenza finita di osservazioni $e^{n} \in E^{*}$, si definisce il cilindro

$$[e^{n}] = \{ \sigma \in E^{\omega} : \sigma|n = e^{n} \}$$

l'insieme delle sequenze infinite che hanno e^n come prefisso. L'insieme di tutti i cilindri genera una topologia su E^ω (la topologia prodotto sull'insieme di Baire E^ω, dove E è discreto). Si trasporta questa topologia a Ω tramite la funzione di verità: per ogni insieme di evidenza finita e^n, si definisce

$$O(e^n) = \{ h \in \Omega : V(h, \sigma) = 1 \text{ per qualche } \sigma \in [e^n] \}$$

e si osserva che la topologia generata dagli insiemi O(e^n) (al variare di e^n ∈ E^*) è la topologia su Ω indotta dalla struttura dell'evidenza disponibile. In questa topologia, un insieme $H \subseteq \Omega$ è *aperto* se per ogni $h \in H$ esiste una sequenza di evidenza finita e^n tale che ogni ipotesi compatibile con e^n appartiene a H — ovvero, se l'appartenenza a H è *verificabile in tempo finito*.

Questa connessione tra apertura topologica e verificabilità finita è il cuore del contributo kellyiano alla filosofia della scienza: le proprietà *verificabili* di un'ipotesi sono precisamente le proprietà aperte nella topologia dell'evidenza. Le proprietà *falsificabili* (nel senso popperiano) sono le proprietà chiuse. Le proprietà che sono né verificabili né falsificabili occupano una posizione topologicamente più complessa: sono insiemi della gerarchia di Borel.

Definizione IV.4 (Gerarchia di Borel): Nella topologia dell'evidenza su Ω, si definisce la gerarchia di Borel per induzione transfinita:

$$\Sigma^0_1 = \{ \text{insiemi aperti} \}$$

$$\Pi^0_1 = \{ \text{insiemi chiusi} \} = \{ \Omega \setminus U : U \in \Sigma^0_1 \}$$

$$\Sigma^0_{\alpha+1} = \{ \text{unioni numerabili di insiemi } \Pi^0_\alpha \}$$

$$\Pi^0_{\alpha+1} = \{ \text{intersezioni numerabili di insiemi } \Sigma^0_\alpha \}$$

I livelli bassi di questa gerarchia hanno interpretazioni epistemiche precise. Σ^0_1 corrisponde alle ipotesi verificabili in tempo finito. Π^0_1 corrisponde alle ipotesi falsificabili (la loro negazione è verificabile). Σ^0_2 corrisponde alle ipotesi che si presentano come limite di successioni di ipotesi falsificate — un'ipotesi è in Σ^0_2 se è verificabile "al limite" ma non necessariamente in un passo finito. Π^0_2 corrisponde alle ipotesi che sono "quasi sempre vere" — possono essere temporaneamente refutate ma si ristabiliscono al limite.

IV.5 Il teorema di caratterizzazione kellyiano

Il risultato più profondo della teoria kellyiana stabilisce una corrispondenza esatta tra la complessità topologica delle ipotesi (misurata dal livello nella gerarchia di Borel) e la

complessità epistemica dell'indagine (misurata dal numero di cambiamenti di congettura necessari):

Teorema IV.2 (Caratterizzazione di Kelly): Sia $H \subseteq \Omega$ un insieme di ipotesi nella topologia dell'evidenza. Allora:

$CC(H) = 0 \quad \text{iff} \quad H \in \Sigma^0_1$ (verificabile in un passo finito)

$CC(H) = 1 \quad \text{iff} \quad H \in \Pi^0_2 \setminus \Sigma^0_1$ (un cambio di mente necessario e sufficiente)

$CC(H) = n \quad \text{iff} \quad H \in \Pi^0_{n+1} \setminus \Sigma^0_n$ (n cambi di mente necessari e sufficienti)

$CC(H) = \omega \quad \text{iff} \quad H \in \Pi^0_\omega \setminus \bigcup_{n<\omega} \Sigma^0_n$ (ω cambi di mente nel caso limite)

Dimostrazione (abbozzo del caso CC(H) = 0): H ha complessità 0 se e soltanto se esiste un metodo M che identifica H senza mai cambiare congettura — ovvero M converge alla risposta corretta h alla prima congettura emessa e non la modifica mai più. Ciò è possibile se e soltanto se, per ogni $h \in H$ e per ogni σ compatibile con h, esiste un prefisso finito $\sigma\restriction n$ tale che ogni ipotesi compatibile con $\sigma\restriction n$ appartiene a H. Ma questa è esattamente la condizione per cui h ha un intorno aperto nella topologia dell'evidenza interamente contenuto in H, ovvero H è un aperto. Dunque $CC(H) = 0$ iff $H \in \Sigma^0_1$.

Dimostrazione (abbozzo del caso CC(H) = 1): Si supponga che H sia in $\Pi^0_2 \setminus \Sigma^0_1$. Allora $H = \bigcap_{n\in\mathbb{N}} U_n$ con U_n aperti, ma H stesso non è aperto. Un metodo M con al più un cambio di congettura deve fare il seguente: emettere una prima congettura h_0 basandosi su un prefisso finito, eventualmente cambiarla una volta a h_1, e poi non cambiarla mai più. La strategia ottimale è la seguente: M congettura la prima ipotesi in H compatibile con l'evidenza finora ricevuta (sfruttando l'apertura degli U_n) e la mantiene finché l'evidenza non la rende incompatibile con H. Poiché H è in Π^0_2, l'incompatibilità è riconoscibile in tempo finito (H è intersezione di aperti). Si verifica formalmente che il numero di cambiamenti richiesto è esattamente 1 e che nessun metodo a 0 cambiamenti può identificare H (poiché H non è aperto). ▮

Questo teorema è uno dei risultati più profondi della filosofia formale della scienza: esso quantifica con precisione il "costo epistemico" dell'indagine su ipotesi di diversa complessità logica, e lo fa mediante un collegamento tra la teoria della misura e la topologia da un lato, e la teoria dell'apprendimento computazionale dall'altro.

IV.6 Il rasoio di Occam e la sua giustificazione formale

Una delle aspirazioni più ambiziose della teoria kellyiana è fornire una *giustificazione puramente formale* del rasoio di Occam — il principio metodologico che prescrive di

preferire le ipotesi più semplici tra quelle compatibili con l'evidenza disponibile. La giustificazione tradizionale di questo principio è pragmatica (le ipotesi semplici sono più maneggevoli) oppure metafisica (la natura è fondamentalmente semplice), entrambe insoddisfacenti dal punto di vista epistemologico.

Kelly propone di fondare il rasoio di Occam sulla nozione di *efficienza epistemica*: un metodo che segue il rasoio di Occam effettua il minor numero possibile di cambiamenti di congettura nel percorso verso la risposta corretta, a parità di garanzie di convergenza.

Per formalizzare questo argomento, Kelly introduce una nozione più raffinata di confronto tra metodi: due metodi M e M' che identificano entrambi una classe H sono comparati non solo in termini del numero massimo di cambiamenti di congettura (complessità nel caso peggiore) ma in termini della loro *dominanza* sulla struttura dell'evidenza.

Definizione IV.5 (Dominanza epistemica): Un metodo M *domina epistemicamente* un metodo M' rispetto a H se per ogni ipotesi $h \in H$ e per ogni sequenza di evidenza σ compatibile con h:

$$mc(M, \sigma) \leq mc(M', \sigma)$$

e la disuguaglianza è stretta per almeno qualche coppia (h, σ). In questo caso si scrive $M < M'$.

Il rasoio di Occam, in questa formalizzazione, si traduce nella seguente asserzione:

Teorema IV.3 (Giustificazione formale del rasoio di Occam, Kelly): Sia $<_O$ l'ordinamento di semplicità su Ω (dove $h <_O h'$ significa che h è più semplice di h'). Il metodo

$$M_O(e^n) = \min_{<_O} \{ h \in \Omega : h \text{ è compatibile con } e^n \}$$

che congettura sempre l'ipotesi più semplice compatibile con l'evidenza, domina epistemicamente qualunque metodo M' che sia *non occamista*, nel senso che esiste qualche sequenza di evidenza su cui M' congettura un'ipotesi meno semplice pur disponendo di alternative più semplici compatibili.

Kelly dimostra questo teorema per una classe ampia di nozioni di semplicità formalizzabili topologicamente: in sostanza, le ipotesi più semplici sono quelle che abitano i livelli più bassi della gerarchia di Borel (in particolare i livelli Π^0_2 o Σ^0_1), e il metodo occamista che le preferisce riesce a convergere con meno cambiamenti di congettura su ogni sequenza di evidenza. Il metodo non occamista, che "salta" a ipotesi complesse quando ipotesi semplici sarebbero ancora compatibili,

è costretto a ulteriori cambiamenti di congettura quando l'evidenza successiva forza il ritorno a ipotesi più semplici.

Abbozzo della dimostrazione: Si consideri un problema con due ipotesi h_s (semplice, livello Π^0_2) e h_c (complessa, livello Π^0_3). Sia M_O il metodo occamista che congettura h_s quando h_s è compatibile. Sia M' un metodo che in un certo momento congettura h_c pur essendo h_s compatibile. Si costruisce una sequenza di evidenza σ nel modo seguente: si presenta dapprima evidenza compatibile con entrambe le ipotesi (M' congettura h_c, M_O congettura h_s). Poi si presenta evidenza che distingue: la verità è h_s. M' deve cambiare congettura da h_c a h_s (un cambio aggiuntivo rispetto a M_O). Dunque mc(M', σ) > mc(M_O, σ). ▮

Questo argomento, pur nella sua semplicità schematica, cattura l'essenza della giustificazione kellyiana: il rasoio di Occam è epistemicamente ottimale perché minimizza il numero di revisioni nel percorso verso la verità, e questo è un criterio di eccellenza epistemica puramente formale, non metafisico né pragmatico.

IV.7 Affidabilità, convergenza e verità

Un aspetto centrale della teoria kellyiana che merita trattazione separata è la distinzione tra diversi concetti di *affidabilità* di un metodo d'indagine. Kelly distingue almeno tre nozioni.

Identificazione al limite (Limit Identification, LI): M identifica H al limite se, per ogni h ∈ H e ogni σ compatibile con h, M converge a h su σ. Questa è la nozione goldiana di base. Non dice nulla sull'efficienza della convergenza, né richiede che M sappia *quando* ha raggiunto la risposta corretta.

Identificazione affidabile (Reliable Identification, RI): M identifica H in modo *affidabile* se, oltre a identificare al limite, non converge mai a ipotesi sbagliate: per ogni σ ∈ E^ω, se M converge a h su σ allora V(h, σ) = 1. In altri termini, M non "si ferma su un errore" — ogni congettura stabile è vera. Si nota che RI è una condizione più forte di LI: vi sono classi identificabili al limite ma non identificabili in modo affidabile.

Identificazione con avvertimento (Identification with Mind-Change Bound, MCB): M identifica H con bound k se identifica H al limite e mc(M, σ) ≤ k per ogni h ∈ H e ogni σ compatibile con h. Questo è il concetto che entra nel teorema di caratterizzazione.

La relazione tra queste nozioni e le classi della gerarchia di Borel è la seguente. LI caratterizza le classi di ipotesi *analitiche* (Σ^1_1 nella gerarchia analitica). RI caratterizza le classi Π^0_2 nel caso di ipotesi singole. MCB con bound k caratterizza le classi Π^0_{k+1} come stabilito dal teorema IV.2.

IV.8 La struttura temporale dell'indagine e il confronto con la DEL

Si è accennato nell'introduzione alla differenza strutturale tra la DEL e la teoria di Kelly rispetto alla dimensione temporale. È ora possibile essere più precisi.

Nella DEL, uno stato epistemico è un modello di Kripke (W, {R_i}, V) e un aggiornamento è un'operazione che trasforma questo stato in un nuovo stato. Le trasformazioni sono *azioni puntuali*: ogni aggiornamento è discreto, finitamente specificato (dal modello d'azione), e applicabile in un singolo passo. La sequenza di aggiornamenti è essa stessa finita e controllata: l'analisi riguarda *cosa succede* dopo una sequenza determinata di eventi comunicativi.

Nella teoria di Kelly, invece, la struttura temporale è essenzialmente diversa. Un metodo d'indagine riceve una sequenza *infinita* di osservazioni e deve congetturare *ad ogni passo* una risposta. Non vi è un termine prefissato: l'analisi riguarda il *comportamento al limite* del metodo. La nozione di convergenza è intrinsecamente infinitaria.

Questa differenza si riflette nella struttura matematica dei due approcci. La DEL è una teoria *finito-dimensionale* nel senso che ogni suo risultato è riducibile a un calcolo su strutture finite (grazie ai teoremi di riduzione). La teoria di Kelly è genuinamente infinita: la gerarchia di Borel è una struttura che non ha analoghi finiti non banali, e i teoremi di caratterizzazione fanno riferimento a proprietà di chiusura sotto unioni e intersezioni numerabili.

Vi è tuttavia un punto di contatto profondo: entrambe le teorie modellano un agente razionale che aggiorna le proprie credenze in risposta a nuova informazione. Nella DEL l'agente è modellato come un nodo in una struttura relazionale collettiva, e l'aggiornamento è determinato dalla struttura epistemica globale. In Kelly l'agente è un metodo computazionale isolato che congettura su sequenze di dati. La sintesi tra le due prospettive — la progettazione di metodi d'indagine che operino in ambienti multi-agente, dove la struttura epistemica collettiva evolve sotto la pressione sia di comunicazioni inter-agente sia di osservazioni del mondo esterno — è un programma di ricerca aperto che la letteratura recente ha appena iniziato a esplorare.

Capitolo V — L'Architettura Comune: Convergenze Strutturali, Tensioni Irrisolte e Prospettive

V.1 Due tradizioni, un unico problema

La monografia ha percorso finora due itinerari separati. Il primo, partendo dai fondamenti hintikkaniani della logica epistemica, ha attraversato la svolta plaziana degli annunci pubblici, la generalizzazione gerbrandyana e BMS dei modelli d'azione, fino alla sintesi sistematica di van Ditmarsch, van der Hoek e Kooi: un'architettura formale per la trasformazione degli stati epistemici collettivi sotto la pressione di eventi comunicativi discreti. Il secondo itinerario ha preso le mosse dall'inquadramento

computazionale di Gold, si è addentrato nella teoria kellyiana dell'indagine affidabile con la sua topologia dello spazio delle ipotesi e i suoi teoremi di caratterizzazione, e ha proposto una fondazione formale del rasoio di Occam basata sulla minimizzazione dei cambiamenti di congettura.

Il momento è giunto di porre la domanda che l'accostamento delle due tradizioni implicitamente suscita: in quale rapporto concettuale e formale esse si trovano? Non si tratta di un quesito retorico a cui rispondere con la rassicurante affermazione che "entrambe studiano la conoscenza". La relazione è più profonda e più problematica di quanto non appaia a un primo sguardo, e comporta sia corrispondenze sorprendenti sia divergenze irriducibili che nessuna sintesi superficiale può onestamente oscurare.

V.2 La struttura comune: aggiornamento come funzione di transizione

Il nucleo formale che accomuna le due teorie è la nozione di *trasformazione epistemica*: entrambe modellano un agente (o un insieme di agenti) il cui stato epistemico muta in risposta a stimoli esterni, e si interrogano sulla struttura e sulle proprietà di questa mutazione.

Nella DEL, la trasformazione è catturata dal prodotto di aggiornamento $M \otimes \mathfrak{A}$: dati uno stato epistemico M e un modello d'azione $\mathfrak{A}$, il nuovo stato epistemico è $M \otimes \mathfrak{A}$. Questa è una funzione

$$U : \mathbb{M} \times \mathbb{A} \to \mathbb{M}$$

dove $\mathbb{M}$ è la classe dei modelli epistemici e $\mathbb{A}$ è la classe dei modelli d'azione. La funzione U è parziale (è definita solo quando le precondizioni sono soddisfatte) e si compone naturalmente: $U(U(M, \mathfrak{A}), \mathfrak{B}) = U(M, \mathfrak{A} \circ \mathfrak{B})$ per una composizione opportunamente definita tra modelli d'azione.

Nella teoria di Kelly, la trasformazione è la funzione di congettura del metodo M applicata a sequenze di evidenza crescenti. Per ogni sequenza finita e^n, $M(e^n)$ è la congettura corrente. L'aggiornamento consiste nel passaggio da $M(e^n)$ a $M(e^{n+1})$: dati lo stato corrente (la congettura $M(e^n)$) e la nuova osservazione e_n, il nuovo stato è $M(e^{n+1})$. Anche questo è una funzione di transizione, stavolta sullo spazio delle ipotesi Ω:

$$T : \Omega \times E \to \Omega$$

(con l'avvertenza che T non è puramente markoviano, poiché M può dipendere dall'intera storia e^{n+1} e non solo dalla congettura precedente e dalla nuova osservazione).

La differenza strutturale cruciale è che nella DEL la funzione di transizione è *esternamente specificata* dal modello d'azione (che incorpora la struttura dell'evento

comunicativo), mentre nella teoria di Kelly la funzione di transizione è *il metodo stesso*, oggetto di studio e di valutazione. La DEL chiede: data questa funzione di transizione, quale sarà lo stato epistemico risultante? La teoria di Kelly chiede: quale funzione di transizione dovrebbe adottare un agente razionale che voglia convergere alla verità con il minor numero di revisioni?

V.3 Semantica dei mondi possibili e spazio delle ipotesi: un'isomorfismo parziale

Un secondo punto di contatto strutturale, meno evidente ma concettualmente più profondo, riguarda il rapporto tra i *mondi possibili* della DEL e lo *spazio delle ipotesi* della teoria kellyiana.

In entrambe le teorie, l'ignoranza dell'agente è modellata come incertezza su quale elemento di un insieme sia quello "vero": nella DEL, quale dei mondi in W sia il mondo reale; nella teoria di Kelly, quale delle ipotesi in Ω sia quella vera. In entrambi i casi, l'aggiornamento elimina elementi incompatibili con la nuova evidenza: nella DEL, si elimina tramite la restrizione $M|\varphi$ o il prodotto $M \otimes \mathfrak{A}$; nella teoria di Kelly, un metodo razionale smette di considerare le ipotesi falsificate dall'evidenza ricevuta.

Questo parallelismo suggerisce una traduzione formale. Dato un problema di indagine kellyiano $\langle \Omega, E^{\omega}, V \rangle$, si può costruire un modello epistemico DEL nel modo seguente. Per ogni $n \in \mathbb{N}$ e ogni prefisso $e^n \in E^*$, si definisce il modello

$$M_{e^n} = (\Omega_{e^n}, R, V_{e^n})$$

dove $\Omega_{e^n} = \{ h \in \Omega : h \text{ è compatibile con } e^n \}$ è l'insieme delle ipotesi ancora non falsificate dopo aver ricevuto il prefisso e^n, R è la relazione di accessibilità universale su Ω_{e^n} (un unico agente non conosce ancora nulla di più di quanto stabilito dall'evidenza), e V_{e^n} è la valutazione che codifica le proprietà delle ipotesi. Il passaggio da M_{e^n} a $M_{e^{n+1}}$ quando arriva la nuova osservazione e_n è esattamente una restrizione del modello nel senso plaziano: si elimina ogni ipotesi incompatibile con la nuova osservazione.

In questo senso, la dinamica dell'indagine kellyiana è *una forma* di aggiornamento PAL: ogni osservazione corrisponde all'annuncio di una proposizione che esclude le ipotesi falsificate. La PAL è pertanto, da questo punto di vista, un sottoframework della teoria kellyiana — il framework che descrive un singolo passo di aggiornamento in un processo di indagine potenzialmente infinito.

Tuttavia questa riduzione è parziale e fuorviante se presentata come equivalenza. I punti di differenza irriducibile sono almeno tre, e meritano di essere articolati con precisione.

V.4 Prima divergenza: il tempo e l'infinito

La DEL è una teoria del *presente epistemico*: descrive con eccezionale acutezza la struttura dell'aggiornamento in un singolo passo o in una sequenza finita di passi prefissata. Il suo orizzonte temporale è finito. I teoremi di riduzione, che garantiscono la traducibilità di ogni formula DEL in una formula della logica epistemica statica, esprimono proprio questa finitezza: ogni sequenza di aggiornamenti si "appiattisce" su un singolo stato risultante, calcolabile in anticipo.

La teoria di Kelly è invece intrinsecamente e ineliminabilmente infinitaria. La nozione di convergenza al limite, la gerarchia di Borel, il teorema di caratterizzazione — tutti questi concetti richiedono il riferimento a sequenze infinite di osservazioni e a proprietà che si manifestano "alla fine" del processo, anche quando quella "fine" non è mai raggiunta in tempo finito. Non esiste un analogo kellyiano del teorema di riduzione: la complessità di un problema d'indagine non è riducibile a una proprietà calcolabile in un numero finito di passi.

Questa divergenza ha conseguenze tecniche precise. Nella DEL, il problema di modello checking (dato un modello M, un mondo w e una formula φ, verificare se M, w ⊨ φ) è decidibile in tempo polinomiale nella dimensione di M per il caso proposizionale, e la soddisfacibilità è decidibile in PSPACE. Nella teoria kellyiana, le classi di complessità epistemica interessanti (CC(H) = n per n > 1) corrispondono a classi della gerarchia aritmetica che non sono in generale decidibili: determinare se CC(H) = n richiede, per classi di ipotesi sufficientemente ricche, procedure che non terminano in tempo finito nel caso peggiore.

V.5 Seconda divergenza: singolo agente e multi-agente

La DEL è, nella sua forma più matura, una teoria intrinsecamente *multi-agente*. Il suo oggetto di studio è la struttura epistemica *collettiva* — la distribuzione della conoscenza tra più agenti, la conoscenza comune, la conoscenza distribuita, l'interazione tra le informazioni di agenti diversi. Il prodotto di aggiornamento BMS è definito su modelli con un insieme A di agenti, e le relazioni di indistinguibilità ≈_i del modello d'azione codificano precisamente il fatto che agenti diversi possono interpretare diversamente lo stesso evento comunicativo.

La teoria kellyiana, nella sua formulazione originale, è una teoria del *singolo agente isolato*. Il metodo M è una funzione che mappa sequenze di osservazioni in congetture: non vi è interazione con altri agenti, non vi è trasmissione di informazione, non vi è struttura relazionale collettiva. L'unica fonte di aggiornamento epistemico è la sequenza di osservazioni del mondo esterno, non la comunicazione con altri agenti.

Questo non significa che una teoria kellyiana multi-agente sia impossibile o priva di interesse: al contrario, è un'area di ricerca attiva. Si può definire un problema di indagine multi-agente come una struttura ⟨A, Ω, E^ω, {V_i}_{i∈A}⟩ dove ogni agente i riceve la propria sequenza di osservazioni (potenzialmente diversa da quella degli altri) e produce le proprie congetture, e dove si studia se gli agenti possono convergere alla stessa verità condividendo le proprie congetture. Ma questa estensione introduce una complessità aggiuntiva che la teoria kellyiana originale non affronta.

V.6 Terza divergenza: la natura degli stati epistemici

Una differenza concettuale più sottile, ma forse la più fondamentale, riguarda la natura degli stati epistemici nelle due teorie. Nella DEL, uno stato epistemico è un oggetto relazionalmente ricco: non è semplicemente "l'insieme delle ipotesi compatibili con l'evidenza" ma un modello di Kripke in cui è codificata non solo la conoscenza di primo ordine (quali proposizioni l'agente sa) ma la conoscenza di ogni ordine (cosa l'agente sa di sapere, cosa sa che gli altri sanno, ecc.). La struttura relazionale del modello porta con sé tutta l'architettura delle credenze di ordine superiore.

Nella teoria kellyiana, uno stato epistemico è essenzialmente una congettura — un elemento di Ω — eventualmente accompagnata da un conteggio dei cambiamenti di congettura effettuati finora. Non vi è spazio, nella formulazione standard, per credenze di ordine superiore: il metodo M non ha rappresentazioni interne di ciò che altri agenti credono, né di ciò che crede di sapere sulla propria capacità d'indagine. È una teoria dell'agente epistemicamente "piatto", nel senso che il suo stato interno è una singola ipotesi sul mondo esterno.

Questa differenza riflette diversi obiettivi teorici. La DEL si preoccupa della *struttura* della conoscenza — della sua grammatica formale, delle sue condizioni di coerenza, della sua distribuzione collettiva. La teoria kellyiana si preoccupa della *performance* dell'indagine — di quante revisioni richiede, di quali classi di problemi sono risolvibili, di quali metodi sono ottimali. Sono preoccupazioni complementari piuttosto che in competizione.

V.7 Convergenza profonda: l'epistemologia come teoria dei punti fissi

Nonostante le divergenze ora articolate, esiste un livello più astratto di affinità tra le due teorie che merita di essere messo in luce. Entrambe possono essere lette come teorie dei *punti fissi* in uno spazio epistemico opportunamente definito.

Nella DEL, la conoscenza comune C_G φ è definita come il più piccolo punto fisso dell'operatore

$$X \mapsto E_G(\varphi \wedge X)$$

ovvero come la più piccola soluzione dell'equazione $C_G \varphi = E_G(\varphi \wedge C_G \varphi)$. La semantica relazionale (chiusura transitiva R_G^*) realizza geometricamente questo punto fisso: è il limite dell'iterazione ω-volte dell'operatore "tutti gli agenti in G sanno". In modelli finiti, questo punto fisso è raggiunto in un numero finito di passi; in modelli infiniti, può richiedere iterazioni transfinite.

Nella teoria kellyiana, la convergenza al limite è anch'essa una nozione di punto fisso: il metodo M converge su σ se e soltanto se la sequenza $M(\sigma|0), M(\sigma|1), M(\sigma|2), \ldots$ ammette un valore limite, ovvero se la funzione $n \mapsto M(\sigma|n)$ si stabilizza

definitivamente. La topologia dell'evidenza rende questo preciso: la convergenza è equivalente alla convergenza nella topologia dell'insieme di Baire E^ω.

La corrispondenza formale, pur non essendo un isomorfismo, è dunque la seguente: la conoscenza comune nella DEL e la convergenza al limite nella teoria kellyiana sono entrambe nozioni di punto fisso in sensi topologici correlati. L'epistemologo che volesse costruire una teoria unificata potrebbe prendere come punto di partenza il lemma di Knaster-Tarski sui reticoli completi — il teorema che garantisce l'esistenza di punti fissi per funzioni monotone su reticoli completi — e verificare in quale misura le costruzioni di entrambe le teorie ne siano istanze.

V.8 Il problema dell'aggiornamento con informazione falsa

Un limite condiviso da entrambe le tradizioni — e particolarmente acuto nella DEL — è la difficoltà di modellare aggiornamenti basati su informazione *falsa* o *fuorviante*. La semantica degli annunci pubblici, come si è osservato nel capitolo II, annuncia sempre proposizioni vere: la clausola M, w ⊨ [φ!]ψ richiede che φ sia vera in w come precondizione per l'applicabilità dell'aggiornamento. Se φ è falsa, [φ!]ψ è vacuamente vera — il che equivale a dire che l'annuncio di qualcosa di falso è semplicemente impossibile nella semantica standard.

Questo è un'idealizzazione forte, giustificabile in contesti cooperativi ma inadeguata per la modellazione di inganni, bugie, comunicazioni erronee o self-deception. Estensioni della DEL che abilitano annunci falsi sono state studiate — si parla di *logica dell'inganno* (deceptive announcement logic) e di *logica della credenza basata su evidenza falsa* — ma il loro sviluppo è ancora incompleto, e i teoremi di riduzione perdono la loro forma semplice in questi contesti più ricchi.

Nella teoria kellyiana, il problema è speculare ma diversamente formulato: si assume implicitamente che le osservazioni ricevute dal metodo siano *affidabili* nel senso che non contengono falsi dati. Un metodo che riceve dati corrotti può divergere o convergere a un'ipotesi errata senza che questo rifletta un difetto del metodo stesso. La teoria kellyiana non ha risorse native per distinguere tra fallimento del metodo e fallimento della fonte di dati.

V.9 L'apprendimento multi-agente e le vie di convergenza

La linea di ricerca più promettente per una sintesi delle due tradizioni è l'apprendimento epistemico in ambienti multi-agente. In tale contesto, si considerano n agenti {1, …, n} che osservano il mondo esterno secondo prospettive parzialmente diverse e comunicano le proprie osservazioni o congetture secondo protocolli variabili. Si chiedono: esiste un protocollo di comunicazione tale che gli agenti convergano collettivamente alla verità (nel senso kellyiano), minimizzando i cambiamenti di congettura collettivi? In che misura la struttura epistemica del gruppo — la conoscenza comune, la conoscenza distribuita — accelera o rallenta la convergenza?

Questi problemi possono essere formalizzati come segue. Si ha un insieme di agenti A, uno spazio di ipotesi Ω, e una struttura di evidenza che assegna a ogni agente i una proiezione σ_i : E^ω → E_i^ω dell'evidenza sul suo "canale" privato. Si definisce un *protocollo di comunicazione* come un insieme di regole che specificano quando e cosa ogni agente comunica agli altri. Un metodo multi-agente M^A = {M^A_i}_{i∈A} è un sistema di metodi individuali accoppiati tramite il protocollo di comunicazione.

La domanda kellyiana diventa: esiste M^A tale che per ogni sequenza di evidenza σ, ogni agente i converge a M^A_i(σ↾n) → h (dove h è la vera ipotesi) con un numero totale di cambiamenti di congettura minimale? La risposta dipende sia dalla struttura epistemica iniziale del gruppo (codificata da un modello DEL) sia dalla topologia dello spazio delle ipotesi (codificata dalla struttura kellyiana).

Un risultato parziale in questa direzione può essere formulato come segue:

Proposizione V.1 (Accelerazione epistemica per conoscenza comune): Sia H una classe di ipotesi di complessità CC(H) = k rispetto a un singolo agente. Sia G = {1, …, n} un gruppo di agenti con protocollo di comunicazione tale che ogni osservazione di ogni agente diventa di conoscenza comune in G entro un numero finito di passi dalla sua ricezione. Allora il metodo multi-agente M^G può identificare H con CC(H, G) ≤ ⌈k/n⌉ cambiamenti di congettura collettivi, dove CC(H, G) è la complessità per il gruppo.

L'intuizione è che la condivisione di evidenza (resa possibile dalla comunicazione con conoscenza comune) è equivalente, per il gruppo, alla ricezione di una sequenza di evidenza n volte più ricca rispetto al singolo agente. La struttura di conoscenza comune garantisce che ogni osservazione di qualunque agente è disponibile a tutti in tempo finito, e questo "multi-canale" riduce il numero di cambiamenti di congettura necessari per un fattore proporzionale alla dimensione del gruppo.

Questa proposizione, pur nella sua forma approssimativa, illustra come le risorse della DEL (la conoscenza comune e la sua semantica relazionale) possano accelerare il processo d'indagine kellyiano. La DEL fornisce il *meccanismo* (il protocollo di comunicazione e la struttura epistemica risultante); la teoria di Kelly fornisce il *criterio di valutazione* (il numero di cambiamenti di congettura). La sintesi è metodologicamente naturale.

V.10 Questioni aperte e programmi di ricerca

Alla luce di quanto esposto, si possono individuare almeno cinque direzioni di ricerca aperta che rappresentano i fronti più attivi e più promettenti dell'intersezione tra le due tradizioni.

La prima direzione riguarda la *DEL infinitaria*: è possibile estendere il linguaggio e la semantica della DEL per gestire sequenze infinite di aggiornamenti? In tal caso, i teoremi di riduzione cessano di valere nella forma finita, e si aprono questioni di completezza e decidibilità che richiedono strumenti tecnici diversi — la teoria dei

punti fissi, la logica modale su strutture ordinate, la teoria dei giochi di lunghezza infinita.

La seconda direzione riguarda la *teoria kellyiana con credenze di ordine superiore*: è possibile arricchire il framework kellyiano con stati interni che includano rappresentazioni delle credenze altrui? Un agente kellyiano "di ordine superiore" congettura non solo sulla verità del mondo esterno ma anche sulle congetture degli altri agenti. Questo aggiunge una dimensione strategica (nel senso della teoria dei giochi) alla teoria dell'indagine affidabile.

La terza direzione concerne il *rasoio di Occam in contesti multi-agente*: la giustificazione kellyiana del rasoio di Occam è stata sviluppata per agenti singoli isolati. In un gruppo di agenti che comunicano, il criterio di semplicità potrebbe subire modificazioni: l'ipotesi più semplice per il gruppo non è necessariamente la più semplice per ogni singolo agente. Come si comporta il teorema IV.3 in questo contesto arricchito?

La quarta direzione riguarda le *connessioni con la teoria della misura e la probabilità*: entrambe le teorie hanno legami con la teoria della probabilità soggettiva che meritano di essere sistematizzati. Nella DEL, i modelli di plausibilità possono essere interpretati probabilisticamente, e l'aggiornamento lessicografico ha un'interpretazione come condizionalizzazione bayesiana. Nella teoria kellyiana, i cambiamenti di congettura hanno relazioni con i concetti di martingala e con la teoria dei test statistici. La costruzione di un framework unificato probabilistico-topologico è un problema aperto di grande ambizione.

La quinta direzione, infine, è quella delle *applicazioni ai sistemi informativi multiagente e all'intelligenza artificiale*: la DEL e la teoria kellyiana hanno entrambe applicazioni dirette all'epistemologia degli agenti artificiali — sistemi che devono aggiornare le proprie rappresentazioni del mondo sulla base di input sensoriali e comunicativi. La DEL fornisce un linguaggio formale per specificare i protocolli di comunicazione; la teoria kellyiana fornisce criteri di ottimalità per i metodi di aggiornamento. La loro congiunzione potrebbe costituire la base formale per una teoria dell'agente cognitivo affidabile.

V.11 Una nota metodologica conclusiva

Questa monografia ha percorso un territorio vasto e tecnicamente impegnativo, cercando in ogni momento di mantenere la precisione formale senza sacrificare la profondità filosofica. Vale la pena di concludere con una riflessione sul metodo adottato dalle due tradizioni, che rispecchia concezioni diverse della relazione tra matematica e filosofia.

La DEL è nata da una preoccupazione filosofica — come formalizzare la comunicazione epistemica? — ed è pervenuta a una teoria matematica di notevole profondità e autonomia. Il suo rapporto con la filosofia è rimasto stretto: ogni sviluppo tecnico (i modelli d'azione, l'aggiornamento plausibile, la logica doxastica) risponde a una domanda filosofica precisa (come si modellano le comunicazioni asimmetriche? come si revisano le credenze?). La matematica è al servizio della

filosofia, nel senso che le sue strutture sono costruite apposta per catturare fenomeni epistemici.

La teoria kellyiana procede in direzione opposta: parte da una domanda della filosofia della scienza (quando è razionale un metodo d'indagine?) e scopre che la risposta richiede la mobilitazione di strutture matematiche preesistenti — la topologia di Baire, la gerarchia di Borel, la teoria dei punti fissi — che non erano state concepite per scopi epistemici. La matematica non è costruita ad hoc ma *reclutata*: si scopre che i concetti topologici pertinenti erano già lì, nell'armamentario della matematica pura, in attesa di essere riconosciuti come espressioni formali di concetti epistemologici.

In questa diversità di orientamento metodologico risiede forse il motivo più profondo per accostare le due tradizioni: esse si completano non solo nei loro contenuti teorici ma anche nel loro rapporto con la matematica. La DEL mostra che si può costruire matematica a partire dalla filosofia; la teoria kellyiana mostra che si può trovare filosofia all'interno della matematica. Un'epistemologia formale davvero matura avrebbe bisogno di entrambe le direzioni di traffico.

NOTA BIBLIOGRAFICA

Le opere principali a cui questa monografia fa riferimento sono le seguenti. Per la logica epistemica classica: J. Hintikka, *Knowledge and Belief* (1962); S. Kripke, *Semantical considerations on modal logic* (1963). Per la PAL: J. Plaza, *Logics of public communications* (1989), ristampato in *Synthese* (2007) con un'introduzione storica. Per l'approccio BMS: A. Baltag, L. Moss, S. Solecki, *The logic of public announcements, common knowledge, and private suspicions* (TARK 1998). Per la DEL sistematica: H. van Ditmarsch, W. van der Hoek, B. Kooi, *Dynamic Epistemic Logic* (Springer, 2007). Per la bisimulazione e la completezza: J. Gerbrandy, *Bisimulations on Planet Kripke* (Dissertazione, Amsterdam, 1999). Per la teoria dell'apprendimento: E.M. Gold, *Language identification in the limit* (1967); K. Kelly, *The Logic of Reliable Inquiry* (Oxford, 1996); K. Kelly, *Ockham's Razor, Truth, and Information* (in *Handbook of the Philosophy of Information*, 2008). Per le connessioni con la topologia: O. Schulte e C. Juhl, *Topology as Epistemology* (1996); K. Kelly e C. Glymour, *Why probability does not capture the logic of scientific justification* (1992).

Capitolo VI — Il Teorema di Caratterizzazione per n > 1 e la Revisione delle Credenze AGM

VI.1 La strategia avversariale e la costruzione delle sequenze diagonali

Nel capitolo IV si è dimostrato il teorema di caratterizzazione kellyiano nei casi CC(H) = 0 e CC(H) = 1. Per i casi superiori è necessaria una macchina tecnica più robusta, che si basa sulla costruzione esplicita di sequenze di evidenza avversariali mediante

diagonalizzazione iterata. Questa costruzione — che Putnam nella sua versione proto-kellyiana del problema chiamava "evil demon empirico" — è il cuore combinatorio della teoria.

Prima di procedere, conviene fissare la terminologia. Si lavorerà con il caso più concreto e insieme più generale: quello in cui Ω è un insieme numerabile di ipotesi, E = {0, 1} è un alfabeto binario (ogni osservazione è un bit), e la compatibilità di un'ipotesi h con un prefisso e^n significa che h non è stata *refutata* dall'evidenza e^n — ovvero che esiste almeno una sequenza infinita σ compatibile con h il cui prefisso è e^n. Questa scelta non è restrittiva: ogni framework kellyiano sufficientemente espressivo è riducibile a questa forma.

Definizione VI.1 (Sequenza avversariale): Data una classe di ipotesi H e un metodo M che la identifica, una *sequenza avversariale di ordine k* per M è una sequenza $\sigma \in E^{\omega}$ costruita in k fasi successive, ciascuna delle quali forza M a effettuare almeno un cambio di congettura, nel modo seguente.

Si costruisce σ come limite di una sequenza crescente di prefissi $\sigma_0 \subset \sigma_1 \subset \sigma_2 \subset \dots$ dove ciascun σ_j è un prefisso finito. Il processo di costruzione mantiene un'ipotesi "vera" corrente $h_j \in H$ e un contatore di cambiamenti c_j.

Fase 0: Si pone $\sigma_0 = \varepsilon$ (il prefisso vuoto) e h_0 = qualunque ipotesi in H. Il metodo M, al ricevere σ_0, emette la congettura $M(\sigma_0) = g_0$ (o si astiene). Se si astiene, si estende σ_0 bit per bit finché M non emette una prima congettura g_0. Si registra questo momento come n_0.

Fase $j \to j+1$ (il passo avversariale): Dati σ_j e l'ipotesi corrente h_j, si distinguono due casi. Se $M(\sigma_j) = h_j$, allora M sta congetturando l'ipotesi correntemente "vera": il costruttore avversariale deve costringerlo a cambiare. Per farlo, si sceglie una nuova ipotesi $h_{j+1} \in H$ compatibile con σ_j ma diversa da h_j (tale ipotesi esiste poiché la classe H non ha la proprietà di essere puntualmente verificabile in σ_j, come si mostrerà). Si estende σ_j a σ_{j+1} con una sequenza di bit compatibile con h_{j+1} ma incompatibile con h_j. Questo forza M a cambiare congettura nel passaggio da σ_j a σ_{j+1}.

Se invece $M(\sigma_j) \neq h_j$, allora M sta già congetturando qualcosa di diverso dalla verità corrente: il costruttore può aspettare che M cambi (se non cambia, o converge a qualcosa di sbagliato, il metodo non identifica H), oppure interpretare questo come un cambio già contato.

Il punto cruciale è la possibilità, al passo j, di trovare una nuova ipotesi h_{j+1} compatibile con σ_j. La condizione topologica che garantisce questa possibilità è che l'insieme delle ipotesi compatibili con ogni prefisso finito sia *denso* nella topologia dell'evidenza — il che corrisponde a richiedere che H non sia "puntualmente identificabile" in tempo finito a partire da σ_j.

VI.2 Il caso CC(H) = 2: dimostrazione completa

Si considera ora il caso CC(H) = 2. Si deve dimostrare che:

CC(H) ≤ 2 iff H ∈ Π^0_3

e che

CC(H) = 2 iff H ∈ Π^0_3 ∖ Σ^0_2

La seconda equivalenza segue dalla prima e dal teorema per CC ≤ 1 già dimostrato. Si concentra pertanto sulla prima.

Direzione ⇐ (se H ∈ Π^0_3 allora CC(H) ≤ 2):

Se H ∈ Π^0_3, allora H = ∩{*n*∈ℕ} *U_n con U_n ∈ Σ^0_2. Ogni U_n è a sua volta un'unione numerabile di chiusi: U_n* = ∪{m∈ℕ} C_{n,m} con C_{n,m} ∈ Π^0_1 (chiusi). Si costruisce un metodo M a due cambiamenti di congettura nel modo seguente.

Alla ricezione del prefisso e^k, il metodo M considera l'ipotesi

h^*(e^k) = min_{<_O} { h ∈ H : h è compatibile con e^k }

(la più semplice compatibile con l'evidenza, dove <_O è l'ordinamento di semplicità su Ω). Il metodo emette h^*(e^k) come prima congettura non appena questa è ben definita.

La struttura Π^0_3 garantisce che il processo di convergenza si articola su due "livelli" di oscillazione. Al primo livello, M potrebbe dover cambiare congettura quando l'evidenza forza l'abbandono della prima ipotesi congetturata: questo primo cambio avviene quando l'evidenza entra in un insieme C_{n,m} che esclude la congettura corrente. Al secondo livello, M potrebbe dover cambiare ancora quando la struttura Σ^0_2 del complementare di H si manifesta pienamente. Si dimostra formalmente che non può essere necessario un terzo cambio: la struttura Π^0_3 implica che dopo due aggiustamenti il metodo si stabilizza definitivamente.

Più precisamente: poiché H = ∩*n U_n con U_n ∈ Σ^0_2, ogni U_n* = ∪*m C{n,m} con C*{n,m} chiuso, si ha che la condizione h ∉ H si verifica se e soltanto se ∃ n ∀ m : h ∉ C_{n,m}, ovvero ∃ n : h ∉ U_n. La struttura "∃ ∀" corrisponde esattamente a Σ^0_3 per il complementare di H — quindi H stessa è in Π^0_3. Ogni violazione di H è detectabile in due passi: prima si individua l'indice n (passo Σ^0_1 all'interno di Σ^0_2), poi si verifica che per tutti gli m la condizione C_{n,m} è violata (passo Π^0_1 interno). Questi due passi corrispondono esattamente ai due cambiamenti di congettura del metodo M.

Direzione ⇒ (se CC(H) ≤ 2 allora H ∈ Π^0_3):

Sia M un metodo con cc(M, σ) ≤ 2 per ogni σ compatibile con H. Si deve mostrare che H ∈ Π^0_3.

Per ogni metodo M a al più due cambiamenti, la struttura delle sue traiettorie di congettura è la seguente: M emette g_0, poi eventualmente cambia a g_1, poi eventualmente cambia a g_2, e non cambia più. Poiché M identifica H al limite, g_2 (la congettura finale) è la vera ipotesi h.

Si definisce, per ogni coppia di congetture (g_0, g_1) ∈ Ω × Ω:

S(g_0, g_1) = { σ ∈ E^ω : M fa esattamente le congetture g_0, g_1 e poi stabilizza su h }

Ogni tale insieme S(g_0, g_1) è misurabile nella struttura boreliana di E^ω. Il fatto che M effettui al più due cambiamenti implica che la struttura di S(g_0, g_1) è Π^0_3: si tratta di insiemi definiti da condizioni del tipo "M congettura g_0 per un iniziale segmento finito, poi congettura g_1 per un altro segmento, poi converge", ognuna delle quali è una condizione Π^0_2 o Σ^0_2. La combinazione di due tali condizioni è Π^0_3.

Poiché H = ∪_{(g_0,g_1)} { h ∈ Ω : h è il limite di M su S(g_0,g_1) }, e questa unione è numerabile (Ω è numerabile), si conclude H ∈ Π^0_3. ▌

VI.3 Il caso generale: CC(H) = n

La dimostrazione per il caso generale CC(H) = n procede per induzione su n, usando la struttura ricorsiva della gerarchia di Borel.

Lemma VI.1 (Ascesa nella gerarchia): Sia k ≥ 1. Allora H ∈ Π^0_{k+1} se e soltanto se H può essere scritta come

H = ∩_{n∈ℕ} H_n

dove ogni H_n ∈ Σ^0_k. Equivalentemente, il complementare di H è in Σ^0_{k+1} = ∪_{n∈ℕ} Π^0_k.

Questo è il lemma standard della gerarchia di Borel, che ne caratterizza i livelli pari come chiusure sotto intersezioni numerabili dei livelli dispari precedenti.

Teorema VI.2 (Caratterizzazione completa, Kelly): Per ogni k ≥ 0:

CC(H) ≤ k iff H ∈ Π^0_{k+1}

Dimostrazione per induzione su k:

Il caso k = 0 è il teorema IV.2 per CC = 0: H ∈ Σ^0_1 iff H è verificabile in tempo finito iff CC(H) = 0. Poiché Σ^0_1 = Π^0_1 ∩ Σ^0_1 non è la forma Π^0_{0+1} = Π^0_1 ma Σ^0_1, occorre riformulare la base: CC(H) = 0 iff H ∈ Σ^0_1, che equivale a Π^0_1 per il complementare. Ma conviene riformulare: CC(H) ≤ 0 iff H ∈ Σ^0_1 (aperto), e si nota che Σ^0_1 ⊊ Π^0_1 ⊊ Σ^0_2 ⊊ … è la struttura naturale.

Correggiamo: la forma esatta del teorema (seguendo Kelly) è

CC(H) ≤ k iff H ∈ Π^0_{k+1} per k ≥ 1
CC(H) = 0 iff H ∈ Σ^0_1

con la convenzione Π^0_2 per CC ≤ 1, Π^0_3 per CC ≤ 2, e così via.

Passo induttivo k → k+1: Si assume vero CC(H) ≤ k iff H ∈ Π^0_{k+1}.

⟸: Si supponga H ∈ Π^0_{k+2}. Per il Lemma VI.1, H = ∩*n H_n con H_n* ∈ *Σ^0*{k+1} = ∪*m Π^0_k. Il metodo a k+1 cambiamenti è costruito come segue. Il metodo congettura inizialmente l'ipotesi ottimale compatibile (come nel caso k = 2). Quando l'evidenza forza l'abbandono della congettura corrente, il metodo "sale di un livello" nella struttura Π^0*{k+2}: il passo di salita corrisponde a un cambio di congettura, e per ipotesi induttiva i passi residui sono al più k. Il totale è dunque k+1.

⟹: Sia M un metodo con cc(M, σ) ≤ k+1. Si decompone la struttura delle traiettorie di M come nel caso k = 2: ogni traiettoria con al più k+1 cambiamenti può essere parametrizzata da una sequenza di congetture (g_0, g_1, …, g_{k+1}) ∈ Ω^{k+2}. Per ciascuna tale sequenza, l'insieme delle sequenze di evidenza su cui M effettua esattamente questi cambiamenti in questo ordine è un insieme la cui struttura boreliana è di livello Π^0_{k+2} per ipotesi induttiva (la prima congettura introduce un livello Π^0_1, e ogni cambio successivo aggiunge un livello, portando al totale Π^0_{k+2}). Dunque H ∈ Π^0_{k+2}. ▌

VI.4 Il caso limite: CC(H) = ω

Quando nessun numero finito di cambiamenti è sufficiente ma la convergenza è comunque garantita, si ha CC(H) = ω. La caratterizzazione topologica di questo caso coinvolge il *livello ω della gerarchia di Borel*, che si definisce come

Π^0_ω = ∩_{k<ω} Π^0_k = ∩_{k<ω} Σ^0_k

ovvero l'intersezione di tutti i livelli finiti. In topologia polacca (spazi metrici separabili completi) questo livello è non banale e contiene classi di insiemi che non appartengono ad alcun livello finito.

Teorema VI.3: CC(H) = ω se e soltanto se H ∈ Π^0_ω ∖ ∪{*k*<*ω*} *Π^0*{k+1}.

La dimostrazione è un'estensione transfinita del teorema VI.2: per ogni n, un metodo a n cambiamenti non può identificare H (poiché $H \notin \Pi^0_{n+1}$), ma per ogni n esiste un metodo che identifica H con al più qualche numero di cambiamenti dipendente da n (poiché $H \in \Pi^0_{n+1}$ per ogni n finito). Il metodo che realizza ω cambiamenti congettura iterativamente attraverso i livelli della gerarchia, con ogni livello Π^0_k che richiede potenzialmente un cambio ulteriore.

VI.5 La revisione delle credenze AGM: il terzo vertice

Alchourrón, Gärdenfors e Makinson pubblicano nel 1985 su *Journal of Symbolic Logic* un articolo destinato a fondare matematicamente la teoria della revisione delle credenze razionale. Il loro framework — AGM nell'acronimo standard — parte da una domanda apparentemente semplice: dato un agente che ha un insieme di credenze K e riceve una nuova informazione α che *contraddice* K, come deve razionalmente aggiornare le proprie credenze? La risposta non è banale perché α impone di *abbandonare* qualcosa di K (si parla di *revisione*), a differenza del caso in cui α è coerente con K (dove si ha semplice *espansione*).

Formalmente, le credenze dell'agente sono rappresentate da una *teoria logica* K — un insieme di proposizioni chiuso sotto conseguenza logica: se $K \vdash \varphi$ allora $\varphi \in K$. Le operazioni fondamentali su K sono tre.

Espansione: $K + \varphi = Cn(K \cup \{\varphi\})$, dove Cn è la chiusura sotto conseguenza logica. L'espansione aggiunge φ e tutte le sue conseguenze insieme a K, senza rimuovere nulla. È l'operazione appropriata quando φ è coerente con K.

Contrazione: $K \div \varphi$ è l'insieme di credenze ottenuto rimuovendo φ da K nel modo minimale possibile — si rinuncia a credere φ senza aggiungere nuove credenze. La contrazione è più delicata dell'espansione: bisogna rimuovere abbastanza da K da non poter più derivare φ, ma non di più.

Revisione: $K * \varphi$ è la teoria ottenuta incorporando φ in K anche quando φ contraddice K. La revisione è definita dall'identità di Levi:

$$K * \varphi = (K \div \neg\varphi) + \varphi$$

ovvero: si contrae K rispetto a ¬φ (si rimuove la credenza nella negazione di φ), poi si espande con φ. Questo garantisce che $K * \varphi$ è coerente (contiene φ) e minimizza la perdita di credenze originali.

Definizione VI.2 (Postulati di razionalità AGM): Una funzione di revisione $* : \mathcal{K} \times \mathcal{L} \to \mathcal{K}$ (dove $\mathcal{K}$ è la classe delle teorie logiche e $\mathcal{L}$ il linguaggio) è *razionale nel senso AGM* se soddisfa gli otto postulati seguenti:

(K*1) $K * \varphi$ è una teoria logica (chiusura sotto conseguenza)
(K*2) $\varphi \in K * \varphi$ (la nuova credenza è incorporata)

(K*3) $K * \varphi \subseteq K + \varphi$ (la revisione non aggiunge più di quanto l'espansione aggiungerebbe)
(K*4) Se $\neg\varphi \notin K$, allora $K + \varphi \subseteq K * \varphi$ (se φ è coerente con K, la revisione coincide con l'espansione)
(K*5) $K * \varphi = K_\bot$ solo se φ è contraddittoria (la revisione produce l'insieme inconsistente solo quando si forza una contraddizione)
(K*6) Se $\vdash \varphi \leftrightarrow \psi$, allora $K * \varphi = K * \psi$ (invarianza logica del contenuto della revisione)
(K*7) $K * (\varphi \wedge \psi) \subseteq (K * \varphi) + \psi$ (iterazione della revisione: rivedere rispetto a $\varphi \wedge \psi$ non fa di più che rivedere rispetto a φ poi espandere con ψ)
(K*8) Se $\neg\psi \notin K * \varphi$, allora $(K * \varphi) + \psi \subseteq K * (\varphi \wedge \psi)$ (simmetria del postulato 7)

Il teorema di rappresentazione AGM afferma che queste otto condizioni caratterizzano esattamente le revisioni realizzabili mediante *funzioni di selezione* (selection functions) su insiemi di mondi possibili.

Teorema VI.4 (Rappresentazione AGM): Una funzione di revisione * soddisfa (K*1*)–(K*8) se e soltanto se esiste una *relazione di entrenchment epistemico* $\leq_K$ su $\mathcal{L}$ — un ordine totale che riflette quanto l'agente è "radicato" in ciascuna credenza — tale che

$$K * \varphi = Cn(\{ \psi \in K : \neg\varphi <_K \psi \} \cup \{\varphi\})$$

ovvero la revisione mantiene esattamente quelle credenze di K che sono epistemicamente più solide della negazione di φ, e aggiunge φ.

La relazione di entrenchment $\leq_K$ è l'analogo AGM dell'ordine di plausibilità $\leq_i$ dei modelli DEL: essa ordina le credenze per grado di radicamento, e la revisione segue questa struttura. Il parallelismo formale è evidente.

VI.6 Il triangolo DEL–Kelly–AGM

L'AGM è il terzo vertice di un triangolo teorico il cui corretto inquadramento è essenziale per comprendere il paesaggio dell'epistemologia formale contemporanea. Le tre teorie condividono il medesimo oggetto di studio — la struttura dell'aggiornamento epistemico razionale — ma lo approciano da angolazioni radicalmente diverse che si rispecchiano nelle loro strutture matematiche.

AGM e DEL: Il rapporto tra AGM e DEL è stato oggetto di una letteratura copiosa. La connessione più diretta è attraverso l'aggiornamento lessicografico dei modelli di plausibilità, già discusso nel capitolo III. Katsuno e Mendelzon (1991) hanno mostrato che i postulati AGM corrispondono a condizioni su funzioni di ordine totale su mondi possibili, e Baltag e Smets (2006, 2008) hanno sviluppato sistematicamente la traduzione tra i postulati AGM e le regole di aggiornamento della DEL sui modelli di plausibilità.

La corrispondenza più precisa è la seguente. L'identità di Levi K * φ = (K ÷ ¬φ) + φ corrisponde, sul versante DEL, all'aggiornamento lessicografico ⇑φ seguito dall'aggiornamento condizionalizante su φ. In simboli DEL, dato un modello di plausibilità M e un'ipotesi φ:

M^{*φ} = (M^{⇑φ})|φ

dove M^{⇑φ} è il modello ottenuto riordinando lessicograficamente i mondi di plausibilità secondo φ, e (·)|φ è la restrizione plaziana. Questo schema integra perfettamente l'AGM nella struttura della DEL: la revisione è un aggiornamento in due fasi, la prima delle quali rioringa le plausibilità e la seconda restringe ai mondi in cui φ è vera.

Dal punto di vista assiomatico, van Ditmarsch, van der Hoek e Kooi dimostrano che il sistema DEL sui modelli di plausibilità con l'aggiornamento lessicografico è suono e completo rispetto alla semantica dei modelli di plausibilità, e che il sistema risultante soddisfa esattamente i postulati (K*1*)–(*K*8) di AGM quando tradotti nel linguaggio modale.

AGM e Kelly: Il rapporto tra AGM e la teoria kellyiana è meno diretto ma non meno profondo. Gärdenfors stesso ha notato che i postulati AGM descrivono un singolo passo di revisione senza considerare le conseguenze dinamiche di sequenze di revisioni. In particolare, il postulato (K*7) è noto per generare risultati paradossali in presenza di revisioni iterate: la composizione K * φ * ψ non è in generale determinata dai postulati AGM senza informazioni aggiuntive sulla struttura dell'entrenchment.

Kelly ha mostrato che questa limitazione dell'AGM ha una controparte topologica precisa: i postulati AGM caratterizzano funzioni di revisione che "saltano" nell'ipotesi più semplice compatibile con la nuova evidenza, il che corrisponde topologicamente a operazioni di restrizione all'interno di insiemi Π^0_2. Ogni sequenza di revisioni AGM iterate è pertanto, nella classificazione kellyiana, un processo di complessità al più ω — ma senza alcun controllo sul numero esatto di cambiamenti. La teoria kellyiana fornisce quindi una misura di complessità che l'AGM non possiede.

Reciprocamente, l'AGM fornisce qualcosa che la teoria kellyiana non ha nella sua forma originale: una struttura di *entrenchment* — una gerarchia di resistenza delle credenze alla revisione. Questa struttura, quando inserita nel framework kellyiano, permette di definire nozioni di "semplicità relativa" che vanno oltre la semplice semplicità sintattica (numero di simboli) e catturano la struttura inferenziale delle teorie candidate. Il rasoio di Occam kellyiano diventa così sensibile alla struttura epistemica interna dell'agente — non si preferisce semplicemente l'ipotesi sintatticamante più breve, ma quella più radicata nella struttura attuale di credenze.

DEL e Kelly attraverso l'AGM: Infine, l'AGM fornisce un collegamento mediato tra la DEL e la teoria kellyiana che non è immediatamente visibile dal confronto diretto delle due teorie. La DEL (nella versione con modelli di plausibilità) implementa operazioni che soddisfano i postulati AGM. La teoria kellyiana fornisce criteri di

ottimalità per sequenze di operazioni AGM. Pertanto, il problema kellyiano "quale sequenza di revisioni AGM converge alla verità con il minor numero di cambiamenti?" è precisamente il problema di ottimizzazione della teoria dell'indagine affidabile, ristretto alla classe delle funzioni di revisione AGM. Questo programma di ricerca — la teoria kellyiana applicata alle revisioni AGM — è stato avviato da Schulte (1999) e rappresenta una delle connessioni più produttive tra le tre tradizioni.

VI.7 La rappresentazione unificata: le strutture di Grove

Un formalismo che integra elegantemente tutti e tre i vertici del triangolo è quello introdotto da Adam Grove nel 1988 (*Two Modellings for Theory Change*): le *sfere di Grove* (sphere systems). Una struttura di Grove per un agente con credenze K è una famiglia annidata di insiemi di mondi possibili

$$\mathscr{S} = \{ S_0 \subseteq S_1 \subseteq S_2 \subseteq \ldots \subseteq W \}$$

dove S_0 è l'insieme dei mondi compatibili con K (i mondi "più plausibili"), S_1 aggiunge i mondi "un po' meno plausibili", e così via. La revisione $K * \varphi$ seleziona i mondi in $S_{j(\varphi)} \cap [[\varphi]]$ dove $j(\varphi)$ è il minimo j tale che $S_j \cap [[\varphi]] \neq \varnothing$ — la "sfera più interna" che interseca la proposizione φ.

Le sfere di Grove sono isomorfe agli ordini di plausibilità $\leq_i$ della DEL: un ordine totale su W genera una struttura di sfere, e viceversa. Esse sono anche isomorfe alle relazioni di entrenchment dell'AGM: l'entrenchment di una formula corrisponde al "raggio" della sfera più piccola che la contiene. E nella teoria kellyiana, le sfere di Grove corrispondono a una struttura di *prior* sulla topologia dell'evidenza: le sfere più interne corrispondono alle ipotesi topologicamente più semplici (nei livelli bassi della gerarchia di Borel).

Questa corrispondenza a tre vie — AGM, DEL, Kelly — attraverso le sfere di Grove costituisce il punto di integrazione formale più profondo tra le tre tradizioni, e suggerisce che esse siano, in ultima analisi, tre presentazioni diverse della stessa struttura matematica sottostante: un ordine su possibilità epistemiche che si trasforma in risposta all'evidenza.

APPENDICE A

Completezza dei Sistemi Modali: Tecniche di Modello Canonico e Filtrazione

A.1 Il metodo del modello canonico

La dimostrazione di completezza per sistemi modali — e in particolare per la logica epistemica S5 e i suoi estensioni dinamiche — è uno dei risultati tecnici fondanti dell'intera tradizione. Se ne espone qui la costruzione nella sua forma generale, applicabile poi ai casi specifici della monografia.

Sia Λ un sistema modale normale (contenente tutti i teoremi della logica proposizionale classica, lo schema K : $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$, e la regola di necessitazione: se $\vdash_\Lambda \varphi$ allora $\vdash_\Lambda \Box\varphi$). Il *modello canonico* per Λ è definito come

$M^\Lambda = (W^\Lambda, R^\Lambda, V^\Lambda)$

dove:

$W^\Lambda = \{ \Gamma \subseteq \text{Form} : \Gamma \text{ è } \Lambda\text{-massimamente consistente} \}$

Un insieme $\Gamma \subseteq$ Form è *Λ-massimamente consistente* se (i) $\Gamma \nvdash_\Lambda \bot$ (è consistente) e (ii) per ogni $\varphi \in$ Form, $\varphi \in \Gamma$ oppure $\neg\varphi \in \Gamma$ (è massimale). L'esistenza di tali insiemi è garantita dal lemma di Lindenbaum.

Lemma A.1 (Lindenbaum): Ogni insieme Λ-consistente $\Delta \subseteq$ Form è estendibile a un insieme Λ-massimamente consistente $\Gamma \supseteq \Delta$.

Dimostrazione: Si enumera Form = $\{\varphi_0, \varphi_1, \varphi_2, \ldots\}$ (possibile poiché Form è numerabile). Si definisce per induzione:

$\Gamma_0 = \Delta$
$\Gamma_{n+1} = \Gamma_n \cup \{\varphi_n\}$ se $\Gamma_n \cup \{\varphi_n\}$ è consistente
$\quad = \Gamma_n \cup \{\neg\varphi_n\}$ altrimenti

Si pone $\Gamma = \bigcup_n \Gamma_n$. Si verifica che Γ è consistente (ogni sottoinsieme finito di Γ è consistente per costruzione) e massimale (per ogni φ_n, o $\varphi_n \in \Gamma$ o $\neg\varphi_n \in \Gamma$). ∎

La relazione di accessibilità del modello canonico è

$\Gamma R^\Lambda \Delta \quad \text{iff} \quad \{ \varphi : \Box\varphi \in \Gamma \} \subseteq \Delta$

ovvero Δ è accessibile da Γ se ogni proposizione necessaria secondo Γ è vera in Δ. La valutazione è

$V^{\Lambda}(p) = \{ \Gamma \in W^{\Lambda} : p \in \Gamma \}$

Lemma A.2 (Verità): Per ogni $\Gamma \in W^{\Lambda}$ e ogni $\varphi \in$ Form:

$M^{\Lambda}, \Gamma \models \varphi \quad \text{iff} \quad \varphi \in \Gamma$

Dimostrazione per induzione su φ:

Caso $\varphi = p$: $M^{\Lambda}, \Gamma \models p$ iff $\Gamma \in V^{\Lambda}(p)$ iff $p \in \Gamma$. ✓

Caso $\varphi = \bot$: $M^{\Lambda}, \Gamma \models \bot$ mai. E $\bot \notin \Gamma$ per consistenza di Γ. ✓

Caso $\varphi = \neg\psi$: $M^{\Lambda}, \Gamma \models \neg\psi$ iff $M^{\Lambda}, \Gamma \not\models \psi$ iff $\psi \notin \Gamma$ (per i.h.) iff $\neg\psi \in \Gamma$ (per massimalità). ✓

Caso $\varphi = \psi \wedge \chi$: $M^{\Lambda}, \Gamma \models \psi \wedge \chi$ iff $M^{\Lambda},\Gamma \models \psi$ e $M^{\Lambda},\Gamma \models \chi$ iff $\psi \in \Gamma$ e $\chi \in \Gamma$ (per i.h.) iff $\psi \wedge \chi \in \Gamma$ (poiché Γ è chiuso sotto sostituzione logica). ✓

Caso $\varphi = \Box\psi$:

($\Longrightarrow$) Supponi $M^{\Lambda}, \Gamma \models \Box\psi$. Allora per ogni Δ con $\Gamma R^{\Lambda} \Delta$, $M^{\Lambda}, \Delta \models \psi$, dunque $\psi \in \Delta$ per i.h. Ora si consideri l'insieme $\Box^{-1}(\Gamma) = \{\varphi : \Box\varphi \in \Gamma\}$. Si vuole mostrare $\Box^{-1}(\Gamma) \cup \{\neg\psi\}$ è inconsistente (se fosse consistente, si estenderebbe a un $\Delta \in W^{\Lambda}$ con $\Box^{-1}(\Gamma) \subseteq \Delta$, cioè $\Gamma R^{\Lambda} \Delta$, ma $\neg\psi \in \Delta$, contraddizione). Dunque $\Box^{-1}(\Gamma) \vdash_{\Lambda} \psi$, ovvero $\vdash_{\Lambda} \Box\varphi_1 \wedge \ldots \wedge \Box\varphi_k \rightarrow \psi$ per qualche $\Box\varphi_j \in \Gamma$. Applicando le regole modali: $\vdash_{\Lambda} \Box(\varphi_1 \wedge \ldots \wedge \varphi_k) \rightarrow \Box\psi$, donde $\Box\psi \in \Gamma$.

($\Longleftarrow$) Supponi $\Box\psi \in \Gamma$. Per ogni Δ con $\Gamma R^{\Lambda} \Delta$, si ha $\psi \in \Delta$ per definizione di R^{Λ}, dunque $M^{\Lambda}, \Delta \models \psi$ per i.h., dunque $M^{\Lambda}, \Gamma \models \Box\psi$. ✓ ∎

Corollario A.3 (Completezza per Λ): Se $\not\models_{\Lambda} \varphi$ (φ non è valida nella classe di modelli per Λ), allora $\not\vdash_{\Lambda} \varphi$.

Dimostrazione: Se $\not\models_{\Lambda} \varphi$, esiste un modello M e un mondo w con M, $w \not\models \varphi$, dunque M, $w \models \neg\varphi$, dunque $\{\neg\varphi\}$ è consistente, si estende a $\Gamma \in W^{\Lambda}$ con $\neg\varphi \in \Gamma$. Per il Lemma A.2, $M^{\Lambda}, \Gamma \models \neg\varphi$, dunque $M^{\Lambda}, \Gamma \not\models \varphi$, donde $\not\vdash_{\Lambda} \varphi$ per correttezza. ∎

A.2 Specializzazione a S5: le relazioni di equivalenza nel modello canonico

Per ottenere la completezza di S5 rispetto alla classe dei modelli con relazioni di equivalenza, si deve verificare che la relazione R^{S5} del modello canonico è una relazione di equivalenza.

Riflessività: Si deve mostrare ΓR^{S5}Γ, ovvero {φ : □φ ∈ Γ} ⊆ Γ. Poiché S5 contiene l'assioma T : □φ → φ, se □φ ∈ Γ allora φ ∈ Γ (per chiusura di Γ sotto modus ponens con T). ✓

Transitività: Si supponga ΓR^{S5}Δ e ΔR^{S5}Θ. Si deve mostrare ΓR^{S5}Θ, ovvero: se □φ ∈ Γ allora φ ∈ Θ. Da □φ ∈ Γ e l'assioma 4 : □φ → □□φ, si ottiene □□φ ∈ Γ. Da ΓR^{S5}Δ: □φ ∈ Δ. Da ΔR^{S5}Θ: φ ∈ Θ. ✓

Simmetria: Si supponga ΓR^{S5}Δ. Si deve mostrare ΔR^{S5}Γ, ovvero: se □φ ∈ Δ allora φ ∈ Γ. Equivalentemente, se φ ∉ Γ allora □φ ∉ Δ. Se φ ∉ Γ allora ¬φ ∈ Γ (massimalità). Per l'assioma 5 : ¬□φ → □¬□φ applicato alla forma equivalente ◇φ → □◇φ, da ¬φ ∈ Γ si deriva □¬φ ∈ Γ (applicando 5 a ¬φ nella forma ¬□¬¬φ → □¬□¬¬φ, cioè ◇¬φ non è nella sfera; la derivazione precisa usa: ¬φ ∈ Γ implica □¬φ ∈ Γ per l'assioma 5 nel sistema S5). Da ΓR^{S5}Δ: ¬φ ∈ Δ. Ma allora φ ∉ Δ (per consistenza), dunque □φ ∉ Δ (poiché se □φ ∈ Δ allora φ ∈ Δ per T, contraddizione). ✓ ∎

A.3 Filtrazione e la proprietà del modello finito

Un secondo strumento tecnico di grande utilità nella logica modale è la *filtrazione*, che permette di costruire modelli finiti equivalenti ai modelli canonici infiniti su formule di profondità modale limitata.

Data una formula φ ∈ Form e un modello M = (W, R, V), si definisce la *filtrazione di M rispetto a φ* come il modello

M/φ = (W/φ, R/φ, V/φ)

dove:

W/φ = W / ~_φ (insieme dei mondi quozientato per ~_φ)

con w ~_φ v iff per ogni sottoformula ψ di φ: M, w ⊨ ψ iff M, v ⊨ ψ. Due mondi sono equivalenti se sono indistinguibili da tutte le sottoformule di φ.

La classe quoziente [w]_φ è la classe di equivalenza di w. Le relazioni e la valutazione quoziente sono

[w]_φ R/φ [v]_φ iff ∃ w' ∈ [w]_φ, ∃ v' ∈ [v]_φ : w'Rv'
V/φ(p) = { [w]_φ : p è sottoformula di φ e w ∈ V(p) }

Teorema A.4 (Filtrazione): M, w ⊨ φ iff M/φ, [w]_φ ⊨ φ.

La dimostrazione è per induzione sulla struttura di φ. Il caso degli operatori modali usa la definizione di R/φ in modo che il lemma di verità si preservi. La conseguenza

cruciale è che $|W/\varphi| \leq 2^{|sub(\varphi)|}$, dove $sub(\varphi)$ è l'insieme delle sottoformule di φ: il modello filtrato è *finito* e ha dimensione al più esponenziale nella lunghezza di φ.

Corollario A.5 (Proprietà del modello finito per S5): S5 ha la proprietà del modello finito: se φ è soddisfacibile in un modello S5, allora è soddisfacibile in un modello S5 finito. Questa proprietà implica la decidibilità di S5 (e di PAL, per il teorema di riduzione).

APPENDICE B

La Gerarchia di Borel: Costruzione, Proprietà e Teoremi di Separazione

B.1 Spazi di Baire e topologia prodotto

Il contesto naturale per la teoria kellyiana è lo *spazio di Baire* $\mathbb{N}^{\mathbb{N}}$ — l'insieme di tutte le funzioni $f : \mathbb{N} \to \mathbb{N}$, dotato della topologia prodotto in cui $\mathbb{N}$ è equipaggiato della topologia discreta. I suoi aperti fondamentali sono i *cilindri*

$$[s] = \{ f \in \mathbb{N}^{\mathbb{N}} : s \subseteq f \}$$

dove $s \in \mathbb{N}^{<\omega}$ è una sequenza finita e $s \subseteq f$ significa che s è un prefisso di f. Lo spazio di Baire è un spazio metrico separabile completo (spazio polacco), e questo è essenziale per i teoremi della teoria descrittiva degli insiemi che sottostanno al teorema di caratterizzazione kellyiano.

Nel contesto della teoria kellyiana con alfabeto binario $E = \{0,1\}$, si lavora con lo spazio di Cantor $\{0,1\}^{\omega}$, che ha proprietà topologiche analoghe. Per generalità espositiva si usa $\mathbb{N}^{\mathbb{N}}$.

B.2 La gerarchia di Borel: definizione precisa

La gerarchia di Borel sullo spazio di Baire $\mathbb{N}^{\mathbb{N}}$ è definita per induzione transfinita sugli ordinali. Per convenienza si usa la notazione standard della teoria descrittiva degli insiemi.

Livelli della gerarchia finita:

$\Sigma^0_1 = \{ U \subseteq \mathbb{N}^{\mathbb{N}} : U \text{ è aperto} \}$
$\Pi^0_1 = \{ \mathbb{N}^{\mathbb{N}} \setminus U : U \in \Sigma^0_1 \} = \{ \text{insiemi chiusi} \}$
$\Sigma^0_2 = \{ \bigcup_{n\in\mathbb{N}} F_n : F_n \in \Pi^0_1 \} = F_\sigma$ (unioni numerabili di chiusi)

$\Pi^0_2 = \{ \bigcap_{n\in\mathbb{N}} U_n : U_n \in \Sigma^0_1 \} = G_\delta$ (intersezioni numerabili di aperti)
$\Sigma^0_3 = \{ \bigcup_{n\in\mathbb{N}} G_n : G_n \in \Pi^0_2 \} = G_{\delta\sigma}$
$\Pi^0_3 = \{ \bigcap_{n\in\mathbb{N}} F_n : F_n \in \Sigma^0_2 \} = F_{\sigma\delta}$

Livelli transfiniti (per α ordinale numerabile, $\alpha \geq 1$):

$\Sigma^0_\alpha = \{ \bigcup_{n\in\mathbb{N}} A_n : A_n \in \Pi^0_{\beta_n} \text{ per qualche } \beta_n < \alpha \}$
$\Pi^0_\alpha = \{ \mathbb{N}^\mathbb{N} \setminus A : A \in \Sigma^0_\alpha \}$
$\Delta^0_\alpha = \Sigma^0_\alpha \cap \Pi^0_\alpha$

La gerarchia è *propria* in ciascun livello: per ogni α, esistono insiemi in $\Sigma^0_\alpha \setminus \Pi^0_\alpha$ e in $\Pi^0_\alpha \setminus \Sigma^0_\alpha$. La propietà si dimostra mediante argomenti diagonali.

B.3 Il teorema di separazione di Luzin

Un risultato cruciale per la teoria kellyiana è il *teorema di separazione di Luzin*, che garantisce la separabilità degli insiemi Σ^0_1 disgiunti.

Teorema B.1 (Separazione di Luzin per Σ^0_1): Se $A, B \in \Sigma^0_1$ e $A \cap B = \varnothing$, allora esistono $A', B' \in \Delta^0_1$ (Boreliani di primo livello, ovvero sia aperti che chiusi — *clopen*) tali che $A \subseteq A'$, $B \subseteq B'$ e $A' \cap B' = \varnothing$.

Più in generale, per ogni α:

Teorema B.2 (Separazione per Σ^0_α): Se $A, B \in \Sigma^0_\alpha$ sono disgiunti, esiste $C \in \Delta^0_{\alpha+1}$ tale che $A \subseteq C$ e $C \cap B = \varnothing$.

Questo teorema ha la seguente interpretazione epistemica nella teoria kellyiana: se due ipotesi h_1 e h_2 appartengono alla stessa classe Σ^0_α, allora sono distinguibili da un osservatore con capacità epistemica di livello $\Delta^0_{\alpha+1}$ — ovvero, un metodo con al più α cambiamenti di congettura può alla fine decidere tra h_1 e h_2.

B.4 Insiemi completi e riduzione

Ogni livello della gerarchia di Borel ammette insiemi *completi* — insiemi che sono "i più difficili" al loro livello nel senso che ogni altro insieme al medesimo livello si riduce ad essi mediante funzioni continue.

Definizione B.3: Un insieme $A \subseteq \mathbb{N}^\mathbb{N}$ è *Σ^0_α-completo* se $A \in \Sigma^0_\alpha$ e per ogni $B \in \Sigma^0_\alpha$ esiste una funzione continua $f : \mathbb{N}^\mathbb{N} \to \mathbb{N}^\mathbb{N}$ tale che $B = f^{-1}(A)$. Si dice che B si *riduce continuamente* ad A, scritto $B \leq_W A$ (riduzione di Wadge).

Gli esempi canonici di insiemi completi sono:

$WF = \{ f \in \mathbb{N}^\mathbb{N} : f \text{ codifica un ordinamento ben fondato} \}$ è Π^1_1-completo

IF = { f ∈ ℕ^ℕ : f non codifica un ordinamento ben fondato } è Σ^1_1-completo

Per la gerarchia di Borel finita, gli esempi canonici sono costruiti mediante insiemi di sequenze con proprietà di convergenza. In particolare, per Π^0_2:

A = { σ ∈ {0,1}^ω : σ contiene infiniti 1 } è Π^0_2-completo

(Π^0_2 = G_δ perché A = ∩_n { σ : ∃ k ≥ n : σ_k = 1 }, intersezione di aperti.)

B.5 Il gioco di Banach-Mazur e la determinatezza

Un strumento potente per la teoria kellyiana è il *gioco di Banach-Mazur* associato a un insieme A ⊆ ℕ^ℕ. Il gioco G(A) si svolge tra due giocatori I e II in modo seguente:

I: s_0 s_2 s_4 …
II: s_1 s_3 s_5 …

dove ogni s_i ∈ ℕ^{<ω} è una sequenza finita non vuota. La *giocata* è la sequenza infinita s_0⌢s_1⌢s_2⌢… ∈ ℕ^ℕ. Il giocatore I vince se questa sequenza appartiene ad A; il giocatore II vince altrimenti.

Teorema B.4 (Banach-Mazur): Il giocatore II ha una strategia vincente in G(A) se e soltanto se A è *magro* (di prima categoria, ovvero unione numerabile di insiemi da nessuna parte densi). Il giocatore I ha una strategia vincente in G(A) se e soltanto se A contiene un aperto denso.

L'interpretazione epistemica del gioco di Banach-Mazur è la seguente: il giocatore I è l'agente che cerca di dimostrare che l'ipotesi vera appartiene ad A (posizione epistemica positiva), il giocatore II è il "demone avversariale" che costruisce la sequenza di evidenza cercando di confutare questa tesi. La strategia vincente del giocatore I corrisponde, nella terminologia kellyiana, all'esistenza di un metodo che converge a un'ipotesi in A senza cambiamenti inutili; la strategia vincente del giocatore II corrisponde alla costruzione di una sequenza avversariale che forza cambiamenti di congettura.

Il collegamento con il teorema VI.2 è preciso: la complessità CC(H) = n corrisponde al fatto che il gioco avversariale per H richiede esattamente n "mosse di cambio di congettura" per il giocatore I prima di poter stabilizzarsi — il che corrisponde topologicamente alla profondità n+1 nella gerarchia di Borel.

B.6 Il teorema di Hausdorff e la gerarchia delle differenze

Un raffinamento della gerarchia di Borel particolarmente rilevante per la teoria kellyiana è la *gerarchia delle differenze* di Hausdorff, che classifica gli insiemi secondo il

numero di operazioni di differenza insiemistica necessarie per costruirli a partire dagli aperti.

Definizione B.5: Per ogni ordinale $\alpha \geq 1$ e ogni α-sequenza $(A_\xi)_{\xi<\alpha}$ di aperti con $A_0 \subseteq A_1 \subseteq \ldots \subseteq A_\xi \subseteq \ldots$, si definisce il *D_a-insieme*:

$$D_\alpha(A_\xi)_{\xi<\alpha} = \bigcup_{\xi<\alpha,\ \xi \text{ dispari}} (A_\xi \setminus A_{<\xi})$$

dove $A_{<\xi} = \bigcup_{\eta<\xi} A_\eta$. Gli insiemi D_α sono Boreliani, e per $\alpha = 2$ si ottiene $A_1 \setminus A_0$ — una differenza di aperti.

Teorema B.6 (Hausdorff): Ogni insieme Boreliano $B \in \Delta^0_2 = \Sigma^0_2 \cap \Pi^0_2$ può essere rappresentato come D_α-insieme per qualche α numerabile. Il minimo tale α è la *profondità di Hausdorff* di B.

La profondità di Hausdorff è la misura più fine della complessità epistemica nella teoria kellyiana: essa raffina il numero di cambiamenti di congettura CC(H) distinguendo tra ipotesi dello stesso livello Boreliano sulla base di una struttura ordinale più sottile. Kelley non ha sviluppato esplicitamente questa connessione, ma essa è implicita nel trattamento delle "gerarchie di cambiamenti" nelle sue opere più tarde.

APPENDICE C

Protocolli di Comunicazione e Logica dei Giochi Epistemici

C.1 Giochi epistemici e strategie

Un *gioco epistemico* è una struttura in cui più agenti interagiscono strategicamente in un contesto di informazione parziale, e le strategie ottimali dipendono in modo essenziale dalla struttura epistemica dei giocatori. La DEL fornisce il linguaggio formale per

descrivere tali giochi; la teoria kellyiana fornisce criteri per valutare le strategie di indagine.

Definizione C.1 (Gioco epistemico): Un gioco epistemico è una tupla

$$G = (N, S, \{A_i\}_{i \in N}, u, M_0)$$

dove $N = \{1,\ldots,n\}$ è l'insieme dei giocatori, S è l'insieme degli stati del gioco, A_i è l'insieme delle azioni disponibili per il giocatore i, $u : N \times S \to \mathbb{R}$ è la funzione di utilità, e M_0 è il modello epistemico iniziale che descrive le conoscenze dei giocatori sullo stato iniziale. Una *strategia* per il giocatore i è una funzione $\sigma_i : S_i^* \to A_i$ che mappa la storia osservata del giocatore i (gli stati e le azioni che i ha osservato) in un'azione.

La razionalità comune in G è definita nel linguaggio della DEL come la formula

$$RC = C_N \left(\bigwedge_{i \in N} Rat_i\right)$$

dove Rat_i afferma che il giocatore i massimizza l'utilità attesa rispetto alle proprie credenze. Il teorema di backward induction per i giochi finiti a informazione perfetta afferma che RC implica che i giocatori giocano il profilo di strategie determinato dall'induzione a ritroso.

Teorema C.2 (Backward Induction e DEL): In un gioco finito a informazione perfetta G, il profilo di strategie di backward induction è il solo compatibile con la formula $RC = C_N(\bigwedge_i Rat_i)$ nel modello epistemico standard. Formalmente: per ogni modello M compatibile con G e ogni mondo $w \in W$ in cui $M, w \vDash RC$, il profilo di azioni $(\sigma_1(w), \ldots, \sigma_n(w))$ coincide con il profilo di backward induction.

C.2 Giochi di annuncio e protocolli ottimali

Nel contesto della DEL, un *protocollo di comunicazione* è una sequenza di annunci (o, più in generale, di modelli d'azione) concordata tra gli agenti. Un protocollo è *epistemicamente ottimale* se, al termine della sequenza di annunci, lo stato di conoscenza collettivo è massimale — ovvero, ogni agente sa tutto ciò che è compatibile con la struttura dell'evidenza disponibile.

Formalmente, si definisce il *protocollo di comunicazione completa* per un gruppo G come la sequenza di annunci

$$\mathscr{P}_{G} = (\mathfrak{A}_1, \alpha_1^*), (\mathfrak{A}_2, \alpha_2^*), \ldots, (\mathfrak{A}_k, \alpha_k^*)$$

tale che, per ogni formula $\varphi \in \mathcal{L}_{EL}$ e ogni mondo w del modello iniziale M_0:

$$M_0 \otimes \mathfrak{A}_1 \otimes \ldots \otimes \mathfrak{A}_k, (w, \alpha_1^*, \ldots, \alpha_k^*) \vDash C_G(\varphi \leftrightarrow K_G \varphi)$$

dove $K_G\,\varphi := \bigwedge_{i\in G} K_i\,\varphi$. Ovvero, al termine del protocollo, per ogni proposizione φ della logica epistemica, la sua verità è di conoscenza comune nel gruppo se e soltanto se tutti la conoscono.

Un risultato tecnico di van Ditmarsch, van der Hoek e Kooi garantisce che tale protocollo esiste per ogni modello epistemico finito:

Teorema C.3 (Esistenza del protocollo di comunicazione completa): Per ogni modello epistemico finito M e ogni gruppo $G \subseteq A$, esiste un protocollo finito $\mathscr{P}_G$ che porta alla conoscenza comune di tutti i fatti epistemicamente accessibili al gruppo.

La dimostrazione è costruttiva: si itera la "comunicazione privata a coppie" tra tutti gli agenti del gruppo secondo uno schema a grafo connesso, e si dimostra che dopo un numero di passi proporzionale a $|G|^2$ il gruppo ha raggiunto la conoscenza comune di tutta l'informazione disponibile.

C.3 Il paradosso dei bugiardi e i limiti del framework

Un problema delicato che emerge nella trattazione dei protocolli di comunicazione è quello degli *agenti ingannevoli*. Si è osservato nel capitolo V che la DEL standard non modella annunci falsi. Tuttavia, in molti protocolli di interesse pratico (crittografia, negoziazione, giochi a informazione incompleta) la possibilità di trasmettere informazioni false è essenziale.

La *logica dell'inganno* (deceptive announcement logic, DAL) estende $\mathcal{L}_{PAL}$ con un operatore $[\varphi_!]\psi$ ("l'agente mente affermando φ, e poi ψ vale") la cui semantica è

$$M, w \models [\varphi_!]\psi \quad \text{iff} \quad M, w \models \neg\varphi \;\text{ e }\; M|\neg\varphi \to ?\,\psi$$

La definizione del modello risultante è delicata: se l'agente afferma φ ma φ è falsa in w, gli altri agenti (se razionali) potrebbero aggiornare le proprie credenze in modi dipendenti dalla loro stima della credibilità dell'agente. Il modello risultante dipende dunque da parametri soggettivi che la DEL standard non porta con sé.

Una formalizzazione coerente richiede l'introduzione di *tipi epistemici* nel senso di Harsanyi — strutture gerarchiche di credenze che includono la credenza sulla credibilità altrui — che rappresentano la frontiera aperta più impegnativa nella DEL contemporanea.

APPENDICE D

Connessioni con la Teoria della Computabilità

D.1 Identificazione al limite e grado di Turing

Il framework goldiano ha connessioni profonde con la teoria della computabilità che meritano di essere esplicitate. La nozione di identificazione al limite da sequenze di dati può essere vista come una forma di computazione "debole": un metodo che identifica al limite non calcola la risposta corretta in un numero finito di passi, ma vi converge.

Definizione D.1 (Computazione al limite): Una funzione $f : \mathbb{N} \to \mathbb{N}$ è *computabile al limite* se esiste una funzione ricorsiva primitiva $g : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ tale che per ogni n:

$$f(n) = \lim_{s \to \infty} g(n, s)$$

ovvero $g(n, s)$ converge a $f(n)$ al crescere di s. Le funzioni computabili al limite sono esattamente quelle con grado di Turing $0'$ — il salto del grado 0 (le funzioni ricorsive totali).

Teorema D.2 (Shoenfield, 1959): Una funzione $f : \mathbb{N} \to \mathbb{N}$ è computabile al limite se e soltanto se il suo grafo $\{ \langle n, f(n) \rangle : n \in \mathbb{N} \}$ è un insieme Δ^0_2 nella gerarchia aritmetica.

Questo teorema è il precursore computazionale del teorema di caratterizzazione kellyiano: la classe Δ^0_2 nella gerarchia aritmetica corrisponde a ciò che è "calcolabile al limite in un passo", esattamente come $CC(H) = 1$ corrisponde a Π^0_2 nella gerarchia di Borel. La corrispondenza non è casuale: la gerarchia aritmetica e la gerarchia di Borel sono due manifestazioni della stessa struttura gerarchica, la prima nel contesto della computabilità, la seconda nel contesto della topologia.

D.2 Identificazione con n cambiamenti e la gerarchia aritmetica

Il teorema di Shoenfield si generalizza a tutti i livelli della gerarchia. La *gerarchia aritmetica* è definita come:

$\Sigma^0_1 = \{ A \subseteq \mathbb{N} : A \text{ è ricorsivamente enumerabile} \}$
$\Pi^0_1 = \{ A \subseteq \mathbb{N} : \mathbb{N} \setminus A \text{ è ricorsivamente enumerabile} \}$
$\Sigma^0_{n+1} = \{ A \subseteq \mathbb{N} : A \text{ è r.e. relativo a un oracolo in } \Pi^0_n \}$
$\Pi^0_{n+1} = \{ A \subseteq \mathbb{N} : \mathbb{N} \setminus A \text{ è r.e. relativo a un oracolo in } \Sigma^0_n \}$
$\Delta^0_n = \Sigma^0_n \cap \Pi^0_n$

Il teorema di Shoenfield generalizzato afferma che f è computabile con n cambiamenti di ipotesi (ovvero calcolabile in modo che g(n, s) cambia valore al più n volte in s) se e soltanto se il grafo di f è in Δ^0_{n+1} della gerarchia aritmetica.

Questa connessione chiarisce il significato computazionale del teorema di caratterizzazione kellyiano: le classi Π^0_{k+1} della gerarchia di Borel corrispondono, per le proprietà che ammettono rappresentazione aritmetica, ai livelli Δ^0_{k+1} della gerarchia aritmetica, che caratterizzano le funzioni computabili con al più k cambiamenti nel valore della loro approssimazione al limite.

La teoria dell'indagine affidabile di Kelly è pertanto, da questo punto di vista, una teoria della *complessità della convergenza* — la misura di quante revisioni sono necessarie per approssimare la verità — formulata in termini topologici invece che computazionali, e quindi applicabile a un dominio molto più ampio che quello delle funzioni ricorsive.